Journal of Experimental Psychology

ARTHUR W. MELTON, Editor
HUMAN RESOURCES RESEARCH CENTER
LACKLAND AIR FORCE BASE
SAN ANTONIO, TEXAS

CONSULTING EDITORS

Judson S. Brown, *State University of Iowa*
Frank A. Geldard, *University of Virginia*
James J. Gibson, *Cornell University*
Clarence H. Graham, *Columbia University*
David A. Grant, *University of Wisconsin*
Lloyd G. Humphreys, *Human Resources Research Center*
Donald B. Lindsley, *University of California, Los Angeles*
Neal E. Miller, *Yale University*
Edwin B. Newman, *Harvard University*
Benton J. Underwood, *Northwestern University*
Delos D. Wickens, *Ohio State University*

Lorraine Bouthilet, Managing Editor

JOHN CRERAR LIBRARY
22 1952
DISCARDED

CONTENTS

The Role of Delay of Reinforcement in Determining Reaction Potential: F. A. LOGAN 393

Discrimination of Stimuli Having Two Critical Components When One Component Varies More Frequently Than the Other: A. J. NORTH AND H. B. LEEDY 400

Relative Difficulty of Number, Form, and Color Concepts of a Weigl-Type Problem Using Unsystematic Number Cards: D. A. GRANT AND J. F. CURRAN 408

Exploratory Experiments on the Stimulus Conditions for the Perception of a Visual Surface: J. J. GIBSON AND F. N. DIBBLE 414

Personal and Environmental Determinants of Perception in a Size Constancy Experiment: J. L. SINGER 420

Some Relations Between Stimulus Patterns and Performance in a Continuous Dual-Pursuit Task: P. M. FITTS AND C. W. SIMON 428

The Role of Stimulus Meaning (m) *in Serial Verbal Learning:* C. E. NOBLE 437

Partial Reinforcement in a Gambling Situation: D. J. LEWIS 447

American Psychological Association

Vol. 43 No. 6 June 1952

JOURNAL OF EXPERIMENTAL PSYCHOLOGY

The JOURNAL OF EXPERIMENTAL PSYCHOLOGY is published monthly, two volumes per year, by the American Psychological Association, Inc. The subscription rate per volume is $7.00, or $14.00 annually. Single copies are $1.25. Subscriptions, orders, and business communications should be addressed to the American Psychological Association, Inc., Prince and Lemon Sts., Lancaster, Pa., or 1515 Massachusetts Avenue, N.W., Washington 5, D. C.

This JOURNAL publishes articles which are intended to contribute toward the development of psychology as an experimental science. Experimental work with normal human subjects is favored over work with abnormal or animal subjects. Studies in applied experimental psychology or engineering psychology may be accepted if they have broad implications for experimental psychology or for behavior theory. Normally, articles of a length exceeding 15 printed pages cannot be accepted. However, a single experimental study or an integrated series of studies accomplished simultaneously must be presented in a single article, rather than in a series of articles.

Articles are published in the order of their receipt, except in rare circumstances. Authors who request them are supplied with 50 free reprints without covers. However, under the new APA reprint policy no free reprints will be supplied to authors submitting manuscripts after January 1, 1952. All of the cost of the author's alterations in galley proof is charged to the author, and one-half of the cost of tables, cuts, and formulas.

Immediate publication can be given to articles whose authors assume the full cost of publication. Such articles must undergo the usual review by the Consulting Editors of the JOURNAL. No free reprints are given in the case of these articles.

Address all articles submitted for regular publication or for immediate publication to the editor, Arthur W. Melton, Human Resources Research Center, Lackland Air Force Base, San Antonio, Texas.

The JOURNAL adheres to the conventions concerning bibliographies, citations, preparation of tables and cuts, etc., as described by J. E. Anderson and W. L. Valentine, The preparation of articles for publication in the journals of the American Psychological Association, *Psychol. Bull.*, 1944, 41, 345–376. Special attention is called to the style of numbers and abbreviations (p. 353). The term 'experimenter,' with its various forms, singular, plural, and possessive singular and plural, is abbreviated *E*, *Es*, *E*'s, and *Es*'; analogous abbreviations are used for the terms 'subject' (*S*) and 'observer' (*O*). All manuscripts must be typed double-spaced throughout, including tables, footnotes, references, and titles for figures. Tables, figures, footnotes, references, and titles for figures must be placed on separate pages at the end of the manuscript.

Entered as second-class matter, February 6, 1937, at the post office at Lancaster, Pa., under the Act of March 3, 1879.

Acceptance for mailing at the special rate of postage provided for in paragraph (d-2), Section 34.40, P. L. & R. of 1948, authorized September 11, 1947.

Journal of
Experimental Psychology

VOL. 43, NO. 6 JUNE, 1952

THE ROLE OF DELAY OF REINFORCEMENT IN DETERMINING REACTION POTENTIAL[1]

FRANK A. LOGAN
State University of Iowa[2]

In his *Principles of Behavior*, Hull (3) assumed that the effect of the time of delay of reinforcement and the magnitude of reinforcement in a simple instrumental conditioning situation is to determine the maximum (M_H) to which the learning variable, habit strength ($_sH_R$) will grow. Since habit strength is assumed to be permanently established and never to weaken, certain predictions are possible concerning the effect of changing the delay or the magnitude of reinforcement. Clearly, if an organism, having learned under particular experimental conditions, is later allowed a shorter delay or a greater magnitude of reinforcement following each response, the limit of habit strength is raised. Performance measures will therefore improve as the learning variable increases. But, should the change occur toward a longer delay or a lesser magnitude of reinforcement, the only consistent prediction extricable from the theory is that performance measures will remain unchanged because habit strength is assumed not to decrease and no other constructs are assumed to be a function of these experimental variables.

Recent experiments, by Crespi (1) and Zeaman (7) suggest the inadequacy of this position as far as magnitude of the goal object is concerned. In both experiments, hungry rats were trained to run down a short straight alley to particular amounts of food. Performance was measured in terms of running time, and consistent with the above assumptions, asymptotic performance was superior in a high reward group when compared with a low reward group. When the low reward group was switched to a greater amount of reinforcement, the expected improvement in performance was observed, although the rapidity with which the change occurred questions its explanation on the basis of increased habit strength. Furthermore, running time became immediately slower when animals in the high reward group were offered a smaller amount of food. This is clearly inconsistent with the assumptions outlined above.

[1] Abstracted from a thesis submitted to the Faculty of the Department of Psychology at the State University of Iowa in partial fulfillment of the requirements for the degree of Doctor of Philosophy. The writer is deeply indebted to Professor Kenneth W. Spence.

[2] Now at the Institute of Human Relations, Yale University.

Hull (4) has recently proposed a modification of his postulate system in which this discrepancy between theory and experimental data is eliminated. Magnitude of reinforcement is, according to this revised position, assumed to affect a new incentive construct, K, that acts in a multiplicative manner with habit strength in determining excitatory potential ($_sE_R$). Habit strength will therefore accrue at the same rate under high reward or low reward conditions, but performance will be superior in the high reward group because K is larger. A change in the magnitude of reinforcement will correspondingly change K, and thus mediate a rather abrupt change in the level of performance.

Although data similar to that obtained by Crespi and Zeaman regarding magnitude of reinforcement were not available concerning delay of reinforcement, Hull assumed that the two variables acted in a similar manner to determine performance. His current system (4), therefore, includes another new construct, J, which is assumed to vary with the time of delay of reinforcement and, like K, to multiply habit strength to determine $_sE_R$ and hence performance. Delay of reinforcement should therefore be expected to produce experimental results similar to those obtained with magnitude of reinforcement. The present experiment is an attempt to test this implication.

Method

Apparatus.—The floor plan of the apparatus is shown in Fig. 1. The entire apparatus was painted flat black with the exception of the floor of the response chamber, which was unpainted brass plate. There was a well-shielded 7½-w. bulb placed above each food cup. The apparatus was situated centrally in a darkened room such that the maximum brightness at either illuminated food cup was approximately .004 apparent ft.-candles, while there was little or no visual stimulation from either unlighted portion of the box.

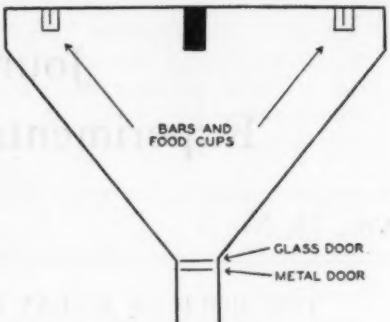

Fig. 1. Floor plan of the apparatus. The walls of the response chamber were 10 in. high; those of the starting alley, 2½ in. high. The food cups were 2 in. above the floor; the bars were ½ in. directly above the food cups. Food cups and bars extended ¾ in. into the response chamber.

The apparatus functioned as follows. When the glass door was raised, a standard electric timer was automatically started. Touching of the bar completed a low amperage contact relay circuit between *S*'s feet on the brass floor and its nose (or paw) on the brass bar. Closure of this circuit led to immediate removal of the bar, stopped the Standard Electric Timer, and fired a Hunter Decade Interval Timer (5). One (or 5) seconds after the response, a .07 gm. pellet of food was dropped automatically into a small food cup situated ½ in. below the extended position of the bar. A 1000-cycle tone was sounded when the pellet was released, and was turned off manually after *S* had obtained the food.

Either bar could function alone as described. When both bars were presented, contact with either bar stopped the Standard Electric Timer, and led to withdrawal of both bars. Food was delivered only in the food cup under the bar to which *S* had responded, and at the interval prescribed by the Hunter Timer fired by that bar.

Subjects.—Forty female hooded rats from the colony maintained by the Department of Psychology of the State University of Iowa served as *Ss*. All *Ss* were naive, and were between 90

and 120 days of age at the beginning of the experiment.

Procedure in preliminary training.—The *S*s were given 7 days habituation to a 24-hr. feeding rhythm. Each rat was fed 7 gm. of dry Purina mash in an individual feeding cage during the same 1-hr. period each day. The *S*s were later run in groups of five rats each, so that the experimental session could be conducted during the 1 hr. immediately preceding the regular feeding time.

All *S*s were given three days of preliminary training in the apparatus, in order to establish the secondary reinforcing properties of the 1000-cycle tone. On these days, the bars were never presented. Each *S* was introduced singly into the response chamber with one bar-light on, and a pellet of mash in that food cup. Thirty seconds after *S* obtained this mash, a pellet was dropped into the same food cup, accompanied by the sound of the 1000-cycle tone. The tone was sounded until *S* had obtained the food. In this manner, four additional pellets were given at 1-min. intervals. The *S* was then removed from the response chamber for 30 min., after which it was given similar training at the opposite food cup. For half of the *S*s, the first training was at the right food cup; for the other half, at the left.

The preliminary training thus consisted of the receipt of 15 pellets of mash from each food cup, each pellet accompanied by the sound of the 1000-cycle tone.

Procedure in training series.—On Day 4 and thereafter, *S*s were run from the starting alley. When *S* was facing toward the response chamber, the metal door was raised and 3 sec. later, the glass door was raised permitting entrance into the response chamber. Reinforcement following the bar response was delayed 1 sec. on one side and 5 sec. on the other. All *S*s were given 160 trials (80 to each bar) under these conditions, the short delay being on the right for a random half of the *S*s. The delays of reinforcement were then reversed for half the *S*s, the other half continuing under the original delay conditions as a control. After 128 additional trials (64 to each bar), the control group *S*s were also subjected to the reversed delays of reinforcement.

For the first four days of the training series, *S*s were given one block of four trials a day; thereafter, two blocks of four trials each were run each day. An average of 5 min. was allowed between trials. The first trial of each block was a free choice run, with both food cups illuminated and both bars presented. The remaining three trials of each block were forced; only one bar was presented and illuminated on these trials. The *S*s could enter the dark side of the response chamber on forced trials, although they soon learned to go directly to the lighted side. The *S*s that chose the short delay bar on the first trial were, on the second trial, half of the time forced to the same bar, and half of the time to the long delay bar. The third and fourth trials were in turn forced in order to equate training to each of the bars. Thus the three forced trials for an *S* with a preference for the short delay bar followed one of the two patterns: SLL or LLS. Since two blocks of four trials each were run per day, each of the two forcing patterns was followed each day, SLL occurring during the first block of four trials on a predetermined half of the days, and during the second on the remaining days. When an *S* chose the long delay bar, the forcing pattern was LSS or SSL, the reverse of that appropriate for an *S* choosing the short delay bar.

On all trials, latency between the opening of the glass door and the bar response was recorded; on each free trial, choice was also obtained. Two measures of differential response strength were thus obtained from each block of four trials: per cent choice on free trials, and latency of response to each bar. The latencies considered were those obtained on the second and fourth trials since, in any pattern of forced trials, *S*s were forced to opposite sides on these two trials. Further, when the pattern was SSLL (or LLSS) each of the compared latencies followed a preceding trial to the same bar, and when the pattern was SLLS (or LSSL), the compared latencies both followed a preceding trial to the opposite bar.

Results

The simple reciprocal of each individual latency was obtained; this measure is hereinafter referred to as *response speed*. The mean response speed curves for the experimental and control groups separately are presented in Fig. 2. The grand mean response speed to the short delay bar for the last six blocks of trials before the reversal for the experimental group alone was .849, while that to the long delay bar was .531. This difference of .318 is significant at beyond the 1% level of confidence ($t = 8.23$). Grand mean response speeds to each bar were calculated for the last six blocks of trials after the reversal of the delays of reinforcement for the experimental group, that to the originally-long-now-short delay bar being .857, and that to the originally-short-now-long delay

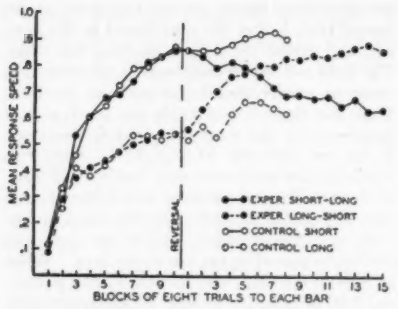

FIG. 2. Mean response speed curves for the experimental and control groups

bar being .626. This difference of .231 is significant at beyond the 1% level of confidence ($t = 6.43$). It is evident that (*a*) *Ss* showed a significantly greater response speed to the short delay bar after original learning, that (*b*) there was a decrease in response speed when the short delay bar was changed to the longer delay of reinforcement, and that (*c*) a corresponding increase in response speed followed the change from a long to a shorter delay of reinforcement.

The performance of the control group, on the other hand, showed no significant change, both response speed curves continuing with a slight increase toward their eventual asymptotes. These final levels of the control group response speeds appear to compare favorably with the post-reversal asymptotes of the response speeds of the experimental group to the appropriate bar.

A corresponding change in behavior following the reversal of the delays of reinforcement is seen in Fig. 3, which shows the per cent choice of the originally short delay bar. The interval 16–84% contains three sigmas on either side of chance expectancy. Since the per cent choice of the originally short delay bar for the last six blocks of trials before the reversal in the delays of reinforcement for the experimental group was 97%, while that during the last six blocks of trials after the reversal was 12.5%, it is obvious that (*a*) *Ss* developed a significant preference for the short delay bar during original learning, and that (*b*) there was a significant change in preference following the reversal of the delays of reinforcement. In contrast, the control group continued their original preference, reaching 100% choice of the short delay bar.

After 144 trials to each bar under the original delay conditions (compared to 80 trials for the experimental group), the control group was also subjected to the reversal of the delays of reinforcement. The effect of this reversal on the response speeds of the control group is shown in Fig. 4. The grand mean response speed to the short delay bar for the last six blocks of trials before the reversal for the control group alone was .905, while that to the long delay bar was .628. This difference of .277 is significant at beyond the 1% level of confidence ($t = 5.57$). Grand mean response speeds to each bar were calculated for the last six blocks of trials after the reversal of the delays of reinforcement for the control group, that to the originally-long-now-short delay bar being

FIG. 3. Choice of the short delay bar by the experimental and control groups

.892, and that to the originally-short-now-long delay bar being .680. This difference of .212 is significant at beyond the 1% level of confidence ($t = 7.46$). It is evident that the control group also showed a decrease in response speed when the short delay bar was changed to the longer delay of reinforcement and a corresponding increase in response speed following the change from a long to a shorter delay of reinforcement.

In order to facilitate comparison of the rates of change of the control group with the experimental group, the post-reversal response speed curves for the experimental group are also presented in Fig. 4, juxtaposed at the point of reversal with the control group. It is noted that the control group response speed curves both begin above the corresponding experimental group curves, suggesting that the experimental group was switched prior to its reaching asymptotic performance under the original delay conditions. The post-reversal curves for the control and experimental groups, however, appear to have the same form, and to approach the same post-reversal asymptotes.

The effect of the reversal of delays of reinforcement on the preference of the control group is shown in Fig. 5,

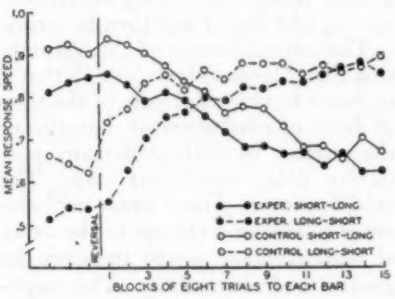

FIG. 4. Post-reversal response speed curves for the experimental and control groups

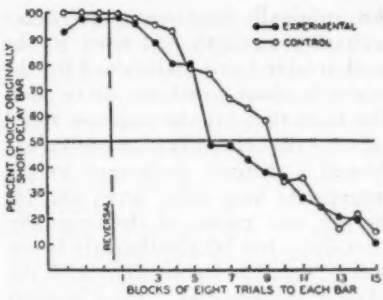

FIG. 5. Post-reversal per cent choice of the short delay bar by the experimental and control groups

where the post-reversal preference of the experimental group is also presented to permit comparison. The control group had attained 100% choice prior to the reversal, while the per cent choice during the last six blocks of trials after the reversal was 16.7. The rate of this change in preference for the control group appears similar to that of the experimental group.

DISCUSSION

It was noted that Hull's original formulation (3), in which the time of delay of reinforcement affected reaction potential by determining the limit to which habit strength could grow, predicted that the excitatory strength to the originally short delay bar would remain at the highest level attained under short delay conditions. From this it follows that the response speed to that bar would not decrease when the delay of reinforcement was changed. Since the originally long delay response speed was expected to increase as a result of additional learning following the advent of the shorter delay of reinforcement, it would be predicted that the per cent choice curve would fall from its prereversal preference for the short delay bar, reaching and staying at or about chance performance when

the originally-long-now-short delay excitatory strength had risen to the level attained and maintained by the originally-short-now-long delay bar. The facts that (*a*) the response speed curve to the originally short delay bar showed a distinct decrement to the appropriate long delay level, and (*b*) the per cent choice of the originally short delay bar fell significantly below chance, stand as inconsistent with the original assumptions. They suggest that, as with magnitude of reinforcement, time of delay of reinforcement should be considered to determine some construct other than habit strength which is at liberty to increase or decrease according to the prevailing conditions of reinforcement.

It should be noted, however, that Hull's newer formulation (**4**) predicts that an appropriate change in behavior will follow a change in the delay of reinforcement in one trial. In this postulate system, the construct J is determined *only* by the time of delay of reinforcement and presumably therefore set at the appropriate value by a single experience of the new reinforcement conditions. The reversal in performance obtained in the present experiment, however, required at least 120 trials to approach completion, which suggests that even Hull's more recent assumption concerning the role of delay of reinforcement in determining reaction potential is inadequate.

Spence (**6**, pp. 272–3) has provided an alternative assumption that J is positively related to the magnitude of the classical conditioned fractional anticipatory goal response.[3] According to such a position, those external and proprioceptive cues coincident with the receipt of the food acquire the capacity to elicit some fractional part of the goal response, $r_g - s_g$. Further, by the principle of stimulus generalization, stimulus events similar to those actually conditioned will also tend to elicit $r_g - s_g$, where the strength of this tendency is assumed to be some positive function of the degree of similarity of the generalized stimulus to the conditioned stimulus. Those stimulus traces present immediately prior to the making of the instrumental response will evoke $r_g - s_g$ because of their similarity to the stimulus traces contiguous with the goal response. It is then but necessary to assume that the shorter the time interval between the occurrence of the generalized stimulus and the conditioned stimulus the greater the similarity of the generalized stimulus trace to the conditioned stimulus trace, in order to derive that $r_g - s_g$ will be larger in an instrumental conditioning situation with a short delay of reinforcement than with a long delay. That is, the stimulus traces immediately preceding the occurrence of the short delay bar response will be more similar to the traces present 1 sec. later when the food is delivered, than will the stimulus traces immediately preceding the occurrence of the long delay bar response to those operant 5 sec. later. The generalized response tendency will therefore be greater under short delay conditions; $r_g - s_g$ and thus J will then be larger.

The critical feature of this assumption that J is related to $r_g - s_g$ is that J is thus a function not only of the time of delay of reinforcement, but also of the number of trials under any particular delay conditions (N). The modification of J (and hence performance) following a change in the delay of reinforcement would therefore be gradual and continuous. This implication is in line with the present findings.

[3] Spence employs a single intervening variable, K, instead of two, J and K. He assumed K is a function of both the magnitude and time of delay of the goal object.

It should be recognized that, in the present experiment, *S*s were responding to two bars, the stimulus properties of which were highly similar. As a result, stimulus generalization from each to the other would be expected to retard the acquisition of the new appropriate value of J. Thus, when the originally short delay bar is followed by a longer delay of reinforcement, the resulting decrement in the tendency for that stimulus to evoke $r_g - s_g$ will generalize; similarly, the increment following a short delay of reinforcement to the originally long delay bar will likewise generalize. In any group of trials, the modification of J that might have obtained because of the new conditions of reinforcement to one bar is reduced by the generalization of an opposite change from the other bar. Thus, the obtained reversal in performance presumably would be slower than would be expected if a single change of delay of reinforcement was administered to *S*.

Summary

1. Forty rats were trained to respond to two bars, one followed by reinforcement after 1 sec., and the other after 5 sec. The *S*s showed significantly greater response speed to the short delay bar, and developed a significant preference for this bar.

2. After 80 trials, the delays of reinforcement were reversed for half of the *S*s. An appropriate reversal in the response speeds and preference followed this change, while no such reversal occurred in the control group while it was continued under the original delay conditions.

3. After an additional 64 trials to each bar, the delays of reinforcement were also reversed for the control group. The resulting reversal in performance was comparable to that obtained from the experimental group.

4. These findings were taken as inconsistent with Hull's original formulation in which time of delay of reinforcement was assumed to determine the maximum to which habit strength could grow.

5. Since the reversal in performance required at least 120 trials, it was suggested that J, in Hull's newer postulate system, should be related to some associative construct, and that this acquisition in the present experiment may have been retarded by stimulus generalization.

(Manuscript received October 5, 1951)

References

1. Crespi, L. P. Quantitative variation of incentive and performance in the white rat. *Amer. J. Psychol.*, 1942, **55**, 467–517.
2. Crespi, L. P. Amount of reinforcement and level of performance. *Psychol. Rev.*, 1944, **51**, 341–357.
3. Hull, C. L. *Principles of behavior.* New York: D. Appleton-Century, 1943.
4. Hull, C. L. Behavior postulates and corollaries—1949. *Psychol. Rev.*, 1950, **57**, 173–180.
5. Hunter, T. A., & Brown, J. S. A decade-type electronic interval-timer. *Amer. J. Psychol.*, 1949, **62**, 570–575.
6. Spence, K. W. Theoretical interpretations of learning. In C. P. Stone, *Comparative psychology.* New York: Prentice-Hall, 1951.
7. Zeaman, D. Response latency as a function of the amount of reinforcement. *J. exp. Psychol.*, 1949, **39**, 466–483.

DISCRIMINATION OF STIMULI HAVING TWO CRITICAL COMPONENTS WHEN ONE COMPONENT VARIES MORE FREQUENTLY THAN THE OTHER[1]

ALVIN J. NORTH AND HERBERT B. LEEDY[2]

Southern Methodist University

A number of experiments have been performed to study the discrimination of stimuli which are "conditional" or "ambivalent" as cues (**2, 3, 4, 7, 8, 9, 10, 11**). In conventional discrimination tasks there usually is discrimination in terms of a single dimension such as size, color, form, etc. But in conditional discrimination tasks success depends on "the capacity to adjust response in accordance with two or more necessary and independent cues" (**7**).

Learning to discriminate stimuli having two (or more) critical components (or aspects) involves special difficulties. For example, suppose the task were to learn to make the Response A to a white circle, B to a black circle, C to a white square, and D to a black square. In this task variations in both color and shape are critical. There is no simple relationship of responses to either stimulus component alone. For example, Responses A and B would be both reinforced and nonreinforced with respect to the attribute of circularity, depending on the color.

In conditional discrimination marked interference during learning is to be expected. First, there would be a tendency for responses to generalize (incorrectly) along each stimulus dimension. Second, there would be generalization in terms of common stimulus elements. In the foregoing example, the response to the *white circle* would tend to generalize to the *white* square, because of the common element of whiteness, and also to the black *circle*, because of the common element of circularity. Conditional discrimination requires response to combinations of discriminated attributes.

Problem

It is to be expected that conditional discrimination learning would depend on the relative frequency with which the critical stimulus components were varied. The purpose of the present experiment, therefore, was to study the discrimination of stimuli having two critical components when one component varies more frequently than the other.

To illustrate, suppose that each of three different forms (circle, triangle, and square) appeared in each of three different colors (red, blue, and green). These two critical components, each having three attributes or values, yield nine unique stimulus patterns. Suppose that the red figures were presented successively for a block of 12 trials, then the blue figures for another block of 12 trials, then the green figures for 12 more trials, and so on, changing the color every 12 trials, but the form every trial. Then form would be varied more frequently than color. For convenience, the less frequently varied component will be designated the "constant" component, since it is held constant within each block of trials. The more frequently varied compo-

[1] This study was supported in part by a grant from the Carnegie Foundation.

[2] Now at the University of Michigan.

nent will be called the "variable" component.

Two questions then arise: (*a*) Which are better discriminated, values of the variable component or values of the constant component? (*b*) Are the responses associated more strongly with values of the variable component or with values of the constant component? The answers to these two questions might well provide guidance in the economy of learning.

METHOD

Subjects.—The *S*s were 51 freshmen women students at two professional nursing schools, 27 from School A and 24 from School B. Their ages varied from 17 to 26 years, with a mean of 19.2. Their ACE test scores had a mean of 103.7 and *SD* of 20.6.

Materials.—The stimuli, of which there were nine, and their appropriate letter responses are shown in Fig. 1. The stimuli and responses were drawn separately in black India ink on 8 × 11-in. sheets of plain white three-ring notebook paper.

The stimuli varied in two critical components: (*a*) *outer* component, and (*b*) *inner* component. Each component had three values, thus yielding nine unique combinations of attributes. It will be noted that stimuli appearing in the same column in Fig. 1 had the same outer component, whereas those in the same row had the same inner component. Each figure thus had its outer component in common with two other stimuli and its inner component in common with two different stimuli. Thus the possibilities for interference due to generalization in terms of common stimulus elements were marked.

In addition to the foregoing materials, which were used in the experiment proper, similar materials were prepared for an initial familiarization phase. These consisted of four simple geometric figures, each having a single letter response, and drawn on separate sheets as described above.

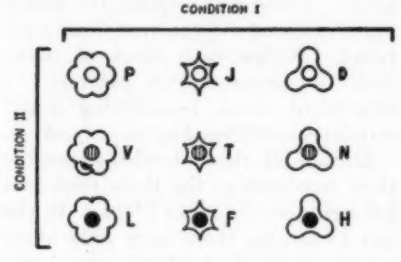

FIG. 1. The stimuli and responses used in the original learning and in the test learning

Presentation of materials.—Group methods for presentation of stimuli were used throughout the experiment. All materials were assembled in stiff-covered notebooks having large rings. Stimuli and responses were presented successively at 4-sec. intervals by manually turning the notebook sheets on a horizontal axis. These operations were timed with the aid of clicks from an electric metronome. The *S*s recorded their responses on specially prepared mimeographed forms having a number of columns with an appropriate space for each response.

Design and procedure.—The experiment consisted of a single session having three phases: (*a*) familiarization, (*b*) original learning (OL), and (*c*) test learning (TL).[2]

The purpose of the familiarization phase was to orient the *S*s to paired associate learning and the recording of responses. The *S*s were told that they would be shown "some patterns and letters" and that they were to learn "which letter goes with each pattern." Then they were shown the familiarization stimuli and responses in order to acquaint them with the manner of presentation. Next they were instructed how to record their responses: Responses were to be recorded in the appropriate space while the stimulus was present and *before* the response sheet was presented. They were encouraged to respond even when not positive of being correct. Otherwise a line was to be drawn in the space. Following these instructions, the four familiarization paired associates were shown for 12 presentations and *S*s recorded their responses.

The OL (original learning) of the paired associates of Fig. 1 followed immediately. In order to compare the learning of the "variable" and the "constant" component, it was necessary to equate their initial discriminability. This was partially accomplished by having the same number of values of each component. More important, any initial differences in the discriminability of the two components were counterbalanced: In Condition I, values of the *outer* component were held constant over blocks of 12 trials, while the inner component varied from trial to trial. In Condition II, values of the *inner* component were held constant over blocks of 12 trials, while the outer component varied from trial to trial. In accordance with this design, *S*s were divided at random into two groups which were subjected to Cond. I

[2] The junior author and an assistant, Mr. John Loper, conducted the experimental sessions.

and II, respectively.[4] The two groups were trained concurrently, one by the junior author and the other by his assistant.

There were 18 blocks of 12 presentations each, or a total of 216 trials, during OL. Between Blocks 9 and 10 there was a short rest period, during which *S*s drew a picture of a person.

Certain rules regarding the order of presentation of stimuli were followed: (*a*) Each value of the variable component appeared once in every three trials, but not twice in succession. (*b*) Each value of the constant component appeared in one of every three blocks, but not in two successive blocks. (*c*) Periodicity was avoided.

The TL (test learning) followed. During OL one component had been varied more frequently than the other. During TL, however, the two components were varied with *equal frequency*, in fact, from trial to trial. Thus the differential effects of the constant and variable components (as established in the OL) on the discrimination of the two-component stimuli could be evaluated.

Accordingly, *S*s of Cond. I and II were jointly tested with 36 presentations of the paired associates of Fig. 1. The outer and inner components were varied from trial to trial. The nine stimuli were repeated in four different orders. Finally, *S*s were asked to make written comments on how they had learned the task.

Results

The responses were analyzed mainly with respect to intrusions. An intrusion was any letter response other than the correct one. If, as happened rarely, an *S* recorded more than one response to a stimulus, only the final response was counted. Extraneous responses, which occurred rarely, were not classified as intrusions.

There were three main kinds of intrusions, defined as follows: (*a*) *Type C* (Constant Type) intrusion was a response belonging to a stimulus identical with the presented stimulus in the component held *constant* (over blocks of 12 trials) in OL. (*b*) *Type V* (Variable Type) intrusion was a response belonging to a stimulus identical with the presented stimulus in the component *varied* (from trial to trial) in OL. (*c*) *Type N* intrusion was a response belonging to a stimulus having *neither* variable nor constant component in common with the presented stimulus.

Two additional points should be made about the foregoing classification. First, Type C intrusions are the same as the *intra*-list intrusions of conventional paired associate experiments; Type V and Type N intrusions taken together correspond to *inter*-list intrusions. Second, on any given trial there was possible one of two Type C, two Type V, or four Type N intrusions. Therefore, the proportionate chance distribution of intrusions of the three types should be 1–1–2.

Analysis of OL.—It will be remembered that OL consisted of 216 trials during which the constant component was varied every 12 trials while the variable component was varied every trial. Thus there were 18 blocks of 12 trials.

Certain background information is of interest. Following a rise in errors on Block 2, there was a gradual downward trend in errors during the series of blocks, with a marked dip in the curve on Block 10, which followed the rest period. During the last three blocks (36 trials) correct responses occurred on 85% of the trials. Failures to respond, or "blockages," showed a concurrent downward trend. Within each block of trials, too, the error curve exhibited a downward trend, manifesting initial interference followed by improvement.

During OL the *S*s tended to confine their responses to the three responses belonging to the given block. In the last 15 blocks, there were 7925 overt responses, of which 91.9% were *intra*-list (correct responses or Type C

[4] One of the 26 Cond. I *S*s was subsequently dropped, thus making the two groups equal in size (25 in each group). This *S* was entirely unable to learn the task within the allotted number of trials, so that her data were of no value.

intrusions) and only 8.1% were *inter*-list (Type V or N intrusions). Apparently, *Ss*' major task during OL was to discriminate among the stimuli of a given block, rather than among stimuli from different blocks.

In Fig. 2 is shown the mean frequency of the several types of intrusions during the series of 18 blocks. The curve for Type C intrusions, despite fluctuations, shows a downward trend, thus demonstrating progressively better discrimination of values of the variable component, which functioned to discriminate among the repertory of three responses belonging to a given block.

The curves for inter-list intrusions (Type V and N) are alike in rising until Block 4 and thereafter gradually declining. Presumably inter-list intrusions were infrequent during the first three blocks because not all the responses had been exhibited until the first two and part of the third block had been completed. It is also probable that not until responses to values of the variable component had been strengthened appreciably would many Type V type intrusions occur.

There was some evidence, however, of generalization on the basis of values of the variable component, particularly as the training progressed. As noted in Fig. 2, Type V intrusions exceeded Type N during the last 12 blocks, even though by chance the frequency of Type N would be double that of Type V intrusions. In the last 15 blocks (excluding the first three, during which the responses first were exhibited) only 2 of the 50 *Ss* had more than twice as many Type N as Type V intrusions, while 44 had just the reverse. These results tend to indicate more generalization in terms of common stimulus elements

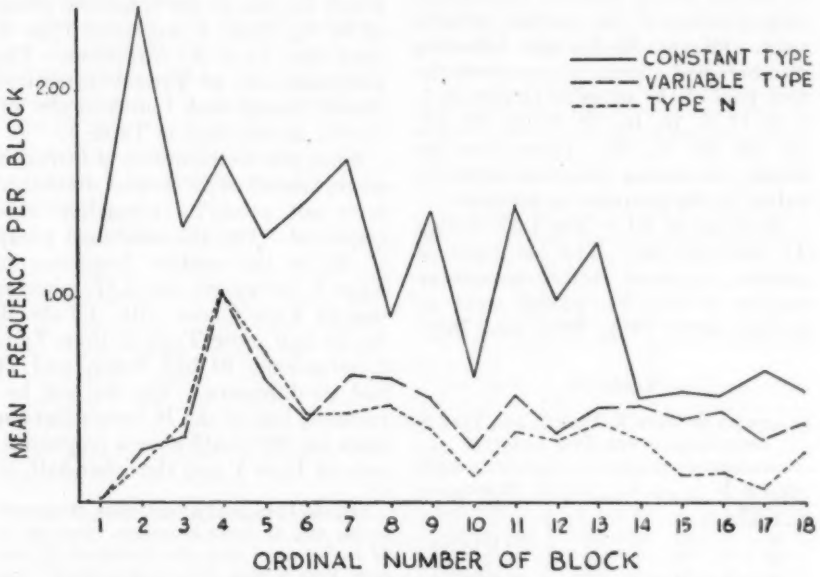
CONSTANT TYPE
VARIABLE TYPE
TYPE N
MEAN FREQUENCY PER BLOCK
2.00
1.00
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18
ORDINAL NUMBER OF BLOCK

FIG. 2. Mean frequency of intrusions per block of 12 trials in the original learning

than along stimulus dimensions, in the case of these materials.

The curves of Fig. 2 also show that Type C exceeded Type V (and also Type N) intrusions in every block. During the last 15 blocks the median (the distributions were skewed) frequency of Type C intrusions was 13.40 whereas that of Type V intrusions was 5.12. Of the 50 *Ss*, 37 had more Type C than Type V intrusions, 9 had fewer, and 4 had no difference. As will be discussed subsequently, the major consequence of this situation was that there was more differential reinforcement of responses to values of the variable component than to values of the constant component.

The data also show that values of the constant component came to be increasingly discriminated. During OL the value of the constant component was changed every 12 trials. An increasing frequency of correct responses on the *first* trial of a series of blocks would indicate discrimination of values of the constant component. Of the 50 *Ss*, the following numbers had correct responses on the first trial of the series of blocks: 0, 1, 0, 9, 17, 8, 16, 16, 20, 33, 22, 25, 17, 29, 29, 29, 31, 39. These data indicate increasing discrimination of values of the constant component.

Analysis of TL.—The task during TL was difficult. The per cents of correct responses during successive fourths of the TL period were as follows: 66%, 72%, 74%, and 76%.

TABLE 1

FREQUENCY OF TYPE V, TYPE C, AND TYPE N INTRUSIONS IN THE TEST LEARNING

Type of Intrusion	Cond. I	Cond. II	Combined Group
V	97	65	162 (54.5%)
C	37	38	75 (25.3%)
N	23	37	60 (20.2%)

TABLE 2

FREQUENCY OF INTRUSIONS IN SUCCESSIVE FOURTHS OF THE TEST LEARNING

Type of Intrusion	Trials			
	1–9	10–18	19–27	28–36
V	41	48	37	36
C	20	18	22	15
N	12	20	12	16

Whereas in the last 36 trials of the OL the per cent of correct responses was 84.9%, the corresponding figure for the TL was 72.3%. Intrusions and blockages, both manifesting interference, occurred on 27.7% of the trials.

Again the analysis of intrusions provided the most valuable information. Intrusions occurred on 16.5% of the trials and accounted for 59.0% of the errors. The major question concerned the relative frequency of Type V and Type C intrusions. Table 1 shows that under both Cond. I and II, and in the combined group of 50 *Ss*, Type V exceeded Type C (and also Type N) intrusions. The preponderance of Type V intrusions existed during each fourth of the TL period, as exhibited in Table 2.

Since the distributions of intrusion scores tended to be skewed, statistical tests not assuming normality were employed. For the combined group of 50 *Ss* the median frequency of Type V intrusions was 2.17, whereas that of Type C was 1.10. Of the 50 *Ss*, 26 had more Type V than Type C intrusions, 10 had fewer, and 14 had no difference.[5] On the null hypothesis, half of the 36 nonambiguous cases (or 18) would have a preponderance of Type V and the other half, of

[5] Of the 14 cases of no difference, 10 were *Ss* having two or fewer intrusions. For the 32 *Ss* having more than two intrusions, 22 had more Type V than Type C intrusions, 8 had fewer, and 2 had no difference.

Type C. Comparing the obtained with the theoretical frequencies, chi square equals 7.11, which is significant beyond the 1% level of confidence.

A comparison of the amount of interference due to generalization along stimulus dimensions and that due to generalization in terms of common stimulus elements is possible for these materials. Hence a comparison of "component" (Types V and C) versus Type N intrusions was made. Hypothetically, Types V and C intrusions would be attributable not only to generalization along stimulus dimensions, but also to generalization in terms of common stimulus elements. On the other hand, Type N intrusions would be attributable only to generalization along stimulus dimensions. On a given trial, there was possible one of four Type N or one of four component intrusions. Hence their probabilities of occurrence were equal, and direct comparison is warranted. During TL 79.8% of the intrusions were of the component type and 20.2% were of Type N. The median frequency of component intrusions was 3.75, whereas that of Type N was .30. Of the 50 *Ss*, 41 had more component than Type N intrusions, 7 had fewer, and 2 had no difference.

Discussion

The results of TL are consistent with the hypothesis that the values of the stimulus component varied most frequently in training are better discriminated than values of the component less frequently varied. The preponderance of Type V intrusions over Type C intrusions in TL supports this hypothesis.

If the training resulted in better discrimination of values of the variable than of the constant component, then why were there more Type C than Type V intrusions during OL? The following explanation is tentatively offered. During the OL the constant component was varied every 12 trials while the variable component was varied every trial. Therefore within each block of trials the only differences among the stimuli were in the variable component. Thus the major task of the *Ss* during a block was to discriminate among the three responses belonging to the block. The extent to which the *Ss* confined their responses to these three responses of a block is shown by the fact that during the last 15 blocks 91.9% of overt responses were intra-list responses, being either correct responses or Type C intrusions. The other 8.1% of overt responses were inter-list intrusions.

What cues, then, did the *Ss* have to discriminate one set of three responses from another? Of course the variations in the constant component provided some differentiation. But it has already been assumed that this source of discrimination among block repertories was of limited effectiveness. The concept of "set" appears to supply an answer. Following Miller (6), it is proposed that response-response associations were formed such that when the response terms of a block were shown (or were given by *S* and reinforced) they tended to place other responses of the repertory in readiness. Some of Miller's results show that when two incompatible series of responses to the same stimuli were practiced in alternation progressively fewer errors occurred. Thus, even in the absence of a specific conditional stimulus to differentiate the lists, inter-list intrusions tended to decrease.

In summary, then, during OL there were more Type C than Type V intrusions because *Ss* tended to confine their responses to the three responses belonging within each block.

How did the OL establish better discrimination of values of the variable than of the constant component? It has already been noted that the *Ss* tended to confine their responses to the repertory of a block and that Type C exceeded Type V intrusions in frequency. The probable answer, then, is that there was more differential reinforcement of responses to values of the variable component than of the constant component.

The argument may be illustrated in the case of Cond. I, in which the outer component was held constant over blocks of 12 trials while the inner component varied every trial. The paired associates in the columns of Fig. 1 were presented in successive blocks of trials. Responses in the same column as the presented stimulus were differentially reinforced with respect to values of the variable component. Responses in the same row as the presented stimulus were differentially reinforced with

respect to values of the constant component. Therefore, since most of the responses during OL were intra-list responses, there was more differential reinforcement of responses to values of the variable than of the constant component.

The grouping of stimuli to be discriminated appears to facilitate discrimination within the groups. Gagné (1) compared the learning of paired associates when similar stimuli were grouped and when they were separated by dissimilar items. In the former case there were more initial intra-list intrusions, but learning proceeded more rapidly.

From the adaptive point of view, it would appear to be more efficient to "pay attention to" those differentiating aspects of a stimulus which vary frequently during a series of trials. In Gestalt terminology, the values of the variable component would become figure and those of the constant component, ground. According to Lashley (5, p. 242),

> In any trial of a training series, only those components of the stimulating situation which are dominant in the organization are associated. Other stimuli which excite the receptors are not associated because the animal is not set to react to them.

The present experiment, however, has demonstrated that *Ss* form associations not only with the dominant (variable) component, but also with the background (constant) component. In Gestalt terminology, then, the results may be summarized under two principles: (*a*) Those critical aspects of the stimulus which are varied most frequently tend to become figure whereas those varied less frequently tend to become less figural, or ground. (*b*) Stronger associations are established to the figure than to the ground aspects of stimuli. Regarding the first principle, the experiment provided that the variable component would become more figural than the constant component. Concerning the second principle, stronger associations would be established to values of the variable than to the constant component. This condition would tend to make for more Type V than Type C intrusions.

Summary

The discrimination of stimuli having two critical components when one component varies more frequently than the other was investigated. Nine paired associates were learned. Each stimulus had an inner and an outer component, each of which had three values. In Cond. I values of the outer component were held constant over blocks of 12 trials while the inner component varied every trial. In Cond. II this relationship was reversed. After the original learning, came the test learning in which both components were varied at equal rates.

The results of the test learning were consistent with the hypothesis that values of the stimulus component varied most frequently in training are better discriminated than values of the component less frequently varied. The better discrimination of values of the variable than of the constant component is attributable to more differential reinforcement of the former during the original learning.

Analysis of intrusions demonstrated that Constant Type were more frequent than Variable Type intrusions in the original learning, but the latter were more frequent during the test learning. The preponderance of Constant Type intrusions in the original learning was attributed not to superior discrimination of values of the constant component, but rather to the limitation of responses to those of a block, which minimized inter-list intrusions.

The hypothesis is offered that frequently varied critical components tend to become more figural than less frequently varied components, and that stronger associations are formed with the former.

(Manuscript received August 28, 1951)

References

1. Gagné, R. M. The effect of sequence of presentation of similar items on the learning of paired associates. *J. exp. Psychol.*, 1950, **40**, 61–73.
2. Galt, W. E. The capacity of the rhesus and cebus monkey and the gibbon to acquire differential response to complex visual stimuli. *Genet. Psychol. Monogr.*, 1939, **21**, 387–457.

3. HARLOW, H. F. Solution of rhesus monkeys of a problem involving the Weigl principle using the matching-from-sample method. *J. comp. Psychol.*, 1943, **35**, 217–227.
4. LASHLEY, K. S. Conditional reactions in the rat. *J. Psychol.*, 1938, **6**, 311–324.
5. LASHLEY, K. S. An examination of the "continuity theory" as applied to discriminative learning. *J. gen. Psychol.*, 1942, **26**, 241–265.
6. MILLER, R. B. Context and stimulus patterning in learning. Unpublished Ph.D. thesis, Yale Univer., 1947.
7. NISSEN, H. W. Ambivalent cues in discriminative behavior of chimpanzees. *J. Psychol.*, 1942, **14**, 3–33.
8. NISSEN, H. W. Analysis of a complex conditional reaction in chimpanzee. *J. comp. physiol. Psychol.*, 1951, **44**, 9–16.
9. NISSEN, H. W., BLUM, J. S., & BLUM, R. A. Analysis of matching behavior in chimpanzee. *J. comp. physiol. Psychol.*, 1948, **41**, 62–74.
10. NOER, M. C., & HARLOW, H. F. Discrimination of ambivalent cue stimuli by macaque monkeys. *J. gen. Psychol.*, 1946, **34**, 165–177.
11. YOUNG, M. L., & HARLOW, H. F. Solution by rhesus monkeys of a problem involving the Weigl principle using the oddity method. *J. comp. Psychol.*, 1943, **35**, 205–217.

RELATIVE DIFFICULTY OF NUMBER, FORM, AND COLOR CONCEPTS OF A WEIGL–TYPE PROBLEM USING UNSYSTEMATIC NUMBER CARDS [1]

DAVID A. GRANT AND JOAN FERRIS CURRAN

University of Wisconsin

The purpose of this study was to investigate the relative difficulty of sorting for color, form, and number in a Weigl-type card sorting problem. The card sorting problem used was the Wisconsin Card Sorting Test (**1, 3, 4, 5**). In an earlier experiment, Grant, Jones, and Tallantis (**5**) found that sorting for color was more difficult than sorting for form or number in the Wisconsin Card Sorting Test (WCST). The greater difficulty in sorting for color and form was due in part to the fact that *Ss* who sorted for number tended to perseverate in their number-sorting habits at stages in the problem when they were supposed to shift from number sorting to form or color sorting. Furthermore, *Ss* learned to sort for number with greater rapidity than they learned to sort for color or form. It was pointed out that the results of the earlier studies seemed to disagree with those of Heidbreder (**6, 7, 8, 9**), but that there were special features in the number concept of the WCST which may have made the number sorting relatively easier in the WCST than it was in Heidbreder's experiments. In the old WCST cards, the number *one* consisted of a single figure, always centered, *two* consisted of two figures along the principal diagonal of the card, *three* figures always formed an equilateral triangle, and *four* figures always formed a square. Thus, *Ss* could sort according to configuration rather than number. In a subsequent research, Grant (**3**) found that an unsystematic configural relationship made for greater difficulty in number sorting especially on the first exposure to the number concept. Verbal reports obtained systematically from *Ss* showed that more than one-third of them claimed to have made use of the systematic configuration. They apparently responded perceptually rather than analytically to the number cards. It was therefore conjectured that repeating the earlier experiment of Grant, Jones, and Tallantis might result in a reorientation of the relative difficulties of number, form, and color sorting, and, in any case, because of the relatively greater purity of the number concept in the revised, unsystematic cards, more light would be shed on the abstract behavior exemplified by the progress of sorting selectively for color or number or form.

[1] This research was supported in part by a grant-in-aid from the Research Committee of the Graduate School of the University of Wisconsin from funds voted by the State Legislature. Acknowledgment is made to the Computing Service of the University for punching IBM cards and making IBM equipment available to the authors for the statistical analysis of the data.

Procedure

The procedure of the present study and the experimental design were essentially identical with that used by Grant, Jones, and Tallantis (**5**). The only differences between the two studies were in cards used and in a slight change in the experimental design so that the differences between *Es* were confounded with the differences among sequences. The WCST has been described earlier (**1, 3, 4, 5**). The test consisted of a pack of 64 response cards, each containing one to four symbols of a single color. Red, green, yellow, and blue colors were used, and

the four symbols used were triangles, stars, crosses, and circles. Each card could, therefore, be sorted for color, form, or the number of symbols. A sorting tray with four double compartments was used. Four stimulus cards were placed in the upper halves of the double compartments. Stimulus cards contained, respectively, one red triangle, two green stars, three yellow crosses, and four blue circles. In the standard procedure, *S* is handed the card, then is instructed to sort them into the lower half of the double compartment under the appropriate stimulus card. He is instructed that he will be told whether each card is sorted "right" or "wrong." Under standard conditions, sorting for color is called "right" until *S* makes ten consecutive correct responses. Then sorting for form is called "right" with no cue to the change other than the shift in "right" and "wrong" called by *E*. After ten consecutive correct responses in stage two, the correct response is changed from form to number. There follow three more stages which repeat the first three, color, form, and number. The *S* must, therefore, shift his sorting behavior five times, each time finding the new principle. There are 24 possible sequences of the three correct concepts in six stages if a concept does not follow itself, and if there are no repetitions until all three responses have been demanded. In the present experiment, all 24 sequences were used to balance out practice and transfer effects. In contrast to the earlier experiment (5) the differences between sequences were confounded with the difference between the two *Es* to eliminate the latter variable from the error estimate, without further complicating the experimental design. Four *Ss* were used with each of the 24 sequences, making 96 *Ss* in all. They were men and women students from elementary psychology classes at the University of Wisconsin. The cards used were the "unsystematic" cards of the previous experiment (3).

Results

Each *S*'s performance at each stage of the experiment was recorded in terms of three scores: (*a*) the number of perseverative errors at each stage except the first, i.e., the number of incorrect responses which would have been correct on the immediately preceding stage except for the initial perseverative error at each stage which presumably served to inform *S* that a shift had taken place; (*b*) the number of *nonperseverative* errors, i.e., the responses that would also have been wrong on the preceding stage of the experiment; and (*c*) the total correct responses, i.e., the number of correct responses made by *S* at each stage of the experiment after he had made his first perseverative error and before he began his criterion sequence of ten successive correct responses.[2] The larger the number of correct responses required by *S* in shifting from one stage to another, the lower was his efficiency in learning the new response.

The results of the experiment are summarized in Fig. 1 in which the total number of correct responses and the average number of perseverative and nonperseverative errors per *S* is given at each stage of the experiment. In the top graph of Fig. 1, it will be seen from the total correct responses required that although number sorting was least efficiently learned in the first stage, it quickly became the most efficient response on all subsequent stages. The next lower figure shows the average number of perseverative errors plotted against stage for the three concepts. By definition, there can be no perseverative errors on the first stage. It will be noted that most of the perseveration takes place on the second stage for *Ss* who are learning to sort for color. Number sorting is hampered by much less perseveration, as is also form sorting. After the second stage, the differences between the form, color, and number sorting become relatively small as far as average perseverative errors are concerned.

The third part of Fig. 1 shows the mean number of perseverative errors on the following stage as a function

[2] The scoring differed from the earlier study by Grant, Jones, and Tallantis (5), but was identical with that used by Grant (3).

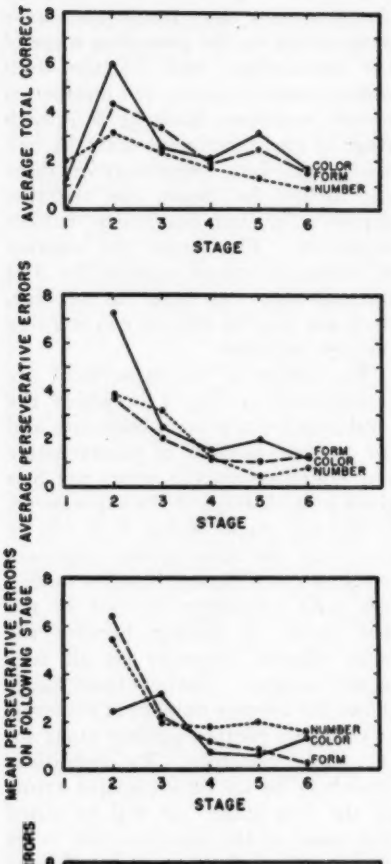

FIG. 1. Scores on successive stages for number, color, and form sorting

of the stage of the experiment. The first perseverative error graph indicated how perseveration hampered the learning of each concept at each stage, but the third graph shows perseveration in a different role. Here is shown the extent to which perseveration of a given sorting response hampers learning to sort for the next concept, whatever it is. In this graph, form and number are shown to perseverate markedly at first; this is shown at Stage 2. The degree of form and number perseveration then decreases sharply, but number sorting tends to perseverate strongly all the way through the experiment and is markedly greater than color and form perseveration on Stages 4, 5, and 6. This same phenomenon stood out in the previous experiment.

The bottom part of Fig. 1 shows the average number of nonperseverative errors at each stage of the experiment for color, form, and number sorting. On the first stage, number sorting produced the most nonperseverative errors. After the first three stages, there are fewer nonperseverative errors made while sorting for number than while sorting for color or form. In this graph, it will be noted that the average number of nonperseverative errors for form was minimal at the first stage of the experiment. The high efficiency of form sorting is also shown in the top part of Fig. 1 where the average total correct required to sort for form on the first stage was zero! These two observations reflect the fact that the number of initial spontaneous sorts for form was higher than that for color and number combined. Of the 96 *S*s, 71 gave an initial spontaneous sort for form, 11 sorted for color, and 14 sorted according to number. In general, Fig. 1 shows

TABLE 1

SUMMARIES OF ANALYSES OF VARIANCE FOR TOTAL CORRECT RESPONSES, PERSEVERATIVE ERRORS, PERSEVERATIVE ERRORS ON FOLLOWING STAGE, AND NONPERSEVERATIVE ERRORS

Source of Variation	*df*	*F*-ratios			
		A. Total Correct Responses	B. Perseverative Errors	C. Perseverative Errors on Following Stage	D. Nonperseverative Errors
(1) Sorting category	2	3.35*	5.27**	3.18*	3.45*
(2) Initial-final sorting	1	11.79***	27.74***	27.74***	37.39***
(3) Initial-final × category	2	1/2.78	1.23	3.14*	10.36***
(4) Sorting order (Stages 1–6)	4	12.97***	32.06***	32.06***	4.98***
(5) Sequences	23	1.75*	2.62**	2.62***	1.46
(6) Pooled error	111	1.96**	1.22	1.22	1.22
(7) Individuals within sequences	72	2.13**	2.07**	2.07	1.42
(8) Error (Individuals within sequences × other factors)	360	(11.14)	(8.99)	(8.99)	(10.52)

Numbers in parentheses are error mean squares.
* Significant at 5% level.
** Significant at 1% level.
*** Significant at 0.1% level.

that sorting for number was most difficult at the beginning of the experiment, but it became most easy of the three sorting responses at the later stages of the experiment. Although the latter was observed in the earlier study, the former result of greater ease of sorting for numbers at the beginning of the experiment had not been observed with the earlier type of systematic response cards.

The statistical analysis of the data presented in Fig. 1 is summarized in Table 1.[3] The entries in Table 1 consist of *F* ratios from the analyses of variance for total correct responses in Column A, perseverative errors in Column B, perseverative errors on the following stage in Column C and nonperseverative errors in Column D. In each column, the error variance estimate is given at the bottom of the column in parentheses, and the mean squares for each row can be obtained by multiplying the *F* ratio by the error estimate. The implication of each variance ratio was described in the earlier paper, so no extended description will be given here.

Attention should be concentrated on the *F*'s in the first three rows in Table 1. The first row of Table 1 shows that the over-all differences due to sorting category, color, number, or form, were significant at the 1% or 5% level for each score that was used. Thus the over-all differences noted in Fig. 1 favoring the number and form sorting over color are found to be significant according to the *F*'s in Table 1. The second line in Table 1 shows that the shift between the first three sorting stages and the second three sorting stages resulted in a significant decrease in total correct responses, perseverative errors and nonperseverative errors.

[3] The fact that the present experiment is virtually an exact repetition of an earlier one (5) affords some opportunity to comment on close statistical comparisons. One notes that, in the present experiment, the error variances are about one-half the order of magnitude of the error variances in the earlier study and that the performance of *S*s was somewhat superior in the present study. (In the earlier experiment, the scoring of perseverative errors and total correct responses was somewhat different, but this has been taken into account in this statement.) The level of difficulty shown by the scores of the present investigation is of the same order of magnitude as that shown in a recent study by Grant (3). Essentially identical cards were used in the last two experiments. In the earlier experiment (5) the cards used were not as well printed as the present set and the symbols were larger. In addition, in the present investigation, the error due to *E*s was confounded with the sequence effect which should tend to reduce the total error variance remaining. It seems likely, however, that the greatest source of reduction of error variance and superior performance of *S*s is due to the different cards that were used in this experiment.

This represents the over-all learning that took place in the second half of the experiment vs. the first half. The third column gives the interaction between the first half of the performance and the second half of the performance and the three sorting categories, color, number, and form. The perseverative errors on the following stage and the nonperseverative errors both show significant interaction. In the perseverative errors on the following stage, the interaction is due largely to the considerable amount of perseveration following number sorting in Stages 4, 5, and 6 as compared with the relatively less degree of number perseveration to be noted in Stages 1, 2, and 3. The interaction found for nonperseverative errors seems to be due primarily to the relative decrease in the number of nonperseverative errors in the form sorting, when one compares Stages 1, 2, and 3 with Stages 4, 5, and 6. It is noteworthy that these interactions proved not to be significant in the earlier study. In general, it might be said that the significance of these interactions arose because of the tremendous improvement in number sorting relative to sorting for color and form which took place progressively throughout the experiment.

Discussion

As compared with the earlier experiment, increasing the purity of the number concept by depriving it of constant configurational aspect increases its difficulty to about the same level as form sorting. Number sorting was especially difficult at the beginning of the problem as had been reported earlier by Heidbreder (**6, 7, 8, 9**). In this respect, the unsystematic number cards of the present experiment give results which agree much more closely with Heidbreder's than did the results of the earlier experiment. The change in the number cards does not reduce the tendency for *Ss* to perseverate in number sorting. This perseveration still hampers learning to sort for color and for form. The tendency for number sorting to perseverate shows little if any reduction after the third stage of the problem; this contrasts sharply with the reduced perseveration of form and color sorting.

Probably the most important finding of this study is the continued perseveration of number sorting. This tendency appeared in the earlier research but is more pronounced in the present experiment. No clue is provided as to the cause of this greater perseveration of number sorting, but one might conjecture that the marked initial difficulty of sorting for number results in a greater degree of overlearning of that response. The fact that certain forms of abstract behavior tend to perseverate more than others has fundamental implications for interpretations of the general process of abstraction. It points up the fact that abstraction is doubtless not a unitary process but has many facets. Simple relative difficulty scores may be unrevealing (**2**), but the tendency of certain responses to perseverate more than others is available as an entering wedge for further experimental analysis of abstract behavior.

Summary

The relative difficulty of sorting for color, form, and number was investigated in a Weigl-type card sorting problem. The experiment essentially repeated an earlier study except that in the present study the constancy of the spatial characteristics of the number concept in the earlier study was reduced so that the number concept was relatively a purer numerical concept. The results and conclusions were as follows:

1. Sorting for color was more difficult over all than sorting for number and form. Number and form appear to be about equally difficult as indicated by the scores of total correct, perseverative errors, and nonperseverative errors.

2. Number sorting was most difficult at the earliest stage of the

experiment but became progressively easier in the later stages. This is in agreement with the findings of Heidbreder.

3. Number sorting was distinguished by marked perseveration as compared with the other abstract responses, particularly in the later stages of the experiment.

This perseveration of number sorting indicates the complexity of the abstraction process and offers a tool for experimental analysis of abstract behavior.

(Manuscript received September 10, 1951)

References

1. Berg, E. A. A simple objective technique for measuring flexibility in thinking. *J. gen. Psychol.*, 1948, **39**, 15–22.
2. Dattman, P. E., & Israel, H. E. The order of dominance among conceptual capacities: an experimental test of Heidbreder's hypothesis. *J. Psychol.*, 1951, **31**, 147–160.
3. Grant, D. A. Perceptual versus analytical responses to the number concept of a Weigl-type card sorting test. *J. exp. Psychol.*, 1951, **41**, 23–29.
4. Grant, D. A., & Berg, E. A. A behavioral analysis of degree of reinforcement and ease of shifting to new responses in a Weigl-type card sorting problem. *J. exp. Psychol.*, 1948, **38**, 404–411.
5. Grant, D. A., Jones, O. E., & Tallantis, B. The relative difficulty of the number, form, and color concepts of a Weigl-type problem. *J. exp. Psychol.*, 1949, **39**, 552–557.
6. Heidbreder, E. The attainment of concepts: VI. Exploratory experiments on conceptualization at perceptual levels. *J. Psychol.*, 1948, **26**, 193–216.
7. Heidbreder, E. The attainment of concepts: VII. Conceptual achievements during card sorting. *J. Psychol.*, 1949, **27**, 3–39.
8. Heidbreder, E., Bensley, M. L., & Ivy, M. The attainment of concepts: IV. Regularities and levels. *J. Psychol.*, 1948, **25**, 299–329.
9. Heidbreder, E., & Overstreet, P. The attainment of concepts: V. Critical features and contexts. *J. Psychol.*, 1948, **26**, 45–69.

EXPLORATORY EXPERIMENTS ON THE STIMULUS CONDITIONS FOR THE PERCEPTION OF A VISUAL SURFACE [1]

JAMES J. GIBSON AND FREDERICK N. DIBBLE

Cornell University

It is possible that the simplest visual perceptions are those of *surface* and *edge* (1). These impressions can be obtained when the eyes are fixated, i.e., without having to consider the effect of the succession of overlapping retinal images normally obtained when the visual environment is scanned. They can probably be explained without reference to the stimulation arising from the posture and the movements of the eyes, head, and body—the factor of equilibrium and orientation in space perception. The vestibular-kinesthetic-tactual complex of stimuli has to be considered, of course, but the visual stimulus might be considered first. This approach to space perception is simplified above all by deferring consideration of the meaning of percepts until their psychophysical basis has been established. The question to which this approach leads, then, is what are the proximal stimuli, in terms of image-variables, for the "elementary impressions" of surface and edge?[2]

A phenomenal surface has the qualities of hardness, distance, slant, and illuminated color (2, p. 368 f.).[3] A phenomenal edge has properties of "two sidedness," or relative distance (a jump in depth from one to the other side), of length, curvature, and direction. An edged or bounded surface has the property of shape or form along with its slant, and size or area along with its distance. For all these properties psychophysical experiments may be possible which will establish the stimulus variables to which they correspond.

According to what will here be called the *texture-hypothesis*, the stimulus for a visual surface is a fully differentiated, sharp, or textured retinal image. This, however, is a crude statement. It can be somewhat refined by the statement that a surface occurs in perception *when the gradients of luminous intensity in the image between small regions of different intensity are maximally steep* (3).

This phraseology implies that a visual surface is a thing. A surface in visual experience is assumed to be something distinct from the filmy or foglike impression obtained when the

[1] Preliminary work on this problem over a period of two years was carried out by the authors. The main experiment to be reported was performed by Dickens Waddell and Walter Carel, to whom especial acknowledgment is due for undertaking to atropinize their own eyes in the interests of the experiment. The research is part of a project carried out under Contract AF41(128)-42 between Cornell University and the Air Force School of Aviation Medicine.

[2] The impression of what the writer has called a "corner," with its qualities of convexity or concavity, will not be discussed here although it also might be considered a basic spatial impression (1, p. 93).

[3] The word *hardness* is not an adequate term for the quality signified. It suggests tactual and kinesthetic meanings which are not intended, such as the softness of fur and the hardness of marble. What is meant is only a kind of visual definiteness which probably goes with the capacity of the eye to accommodate for the surface in question. In this sense, fur and marble are equally hard. Katz described it as something which seemed to stop the eye, as contrasted with the penetrable character of a film or fog (4).

retinal image is homogeneous. Actually, however, it should be considered an open question whether a surface is best studied as an entity or as a variable of visual perception. The question is whether the characteristic quality of hardness, already referred to, is a matter of all-or-nothing, or whether it is at one end of a discriminable dimension with something like "softness" at the other end. The other qualities of surfaceness are surely variables.

The statement of the texture-hypothesis in terms of gradients of luminous intensity suggests the latter interpretation. Since the steepness of such gradients is a variable quantity, the hardness of the impression should be a variable quality. A hard surface would go with the greatest sharpness of texture and a soft film with the least.

A phenomenal edge was characterized as having a definite increase in depth or distance from one side to the other. The texture-hypothesis, by extension, can account for this fact. According to this explanation, depth (along with all other tridimensional qualities) depends in the first instance on the *relative density* of the texture. To this is added the relative disparity of the texture when vision is binocular and the relative displacement of the texture when the head moves (1). The step in relative distance is explained by the step in density, disparity, and motion. The understanding of such a phenomenon is of great practical as well as theoretical importance, for it is approximately what one sees at the edge of a cliff, at the bottom of the windshield when driving a car, and at the line of the cowling when landing an airplane.

It should be noted that this explanation assumes that an edge occurs only between two textured surfaces. If any irreversible jump in phenomenal depth could be demonstrated to occur when the margin in the retinal image is one of intensity or wave length only, the hypothesis would fail to that extent.

The Experiments

The adequacy of the texture-hypothesis may be tested in a general way by making controlled observations with devices which produce the retinal images supposedly necessary for phenomenal surfaces and edges. The experiments to be reported are exploratory in this sense. They were performed with only a few *O*s and invite repetition by others.

Appearance of an extended surface under variable illumination.—A 20-ft. stretch of wall made of a coarse-textured plasterboard and painted a light gray could be made to fill nearly the whole of *O*'s visual field if he were seated about 3 ft. in front of it. The wall was illuminated by a bank of ceiling lights controlled by a rheostat. Even though this device was much simpler than that of Metzger (6), a number of his observations could be checked with it.

At full illumination, the experience was not only surfacelike in the sense of hard, as Metzger has reported, but also visibly textured in a way different from plywood, cloth, or concrete. The texture of this type of plasterboard could be identified and described. In addition to hardness, the perception also involved a definite distance, a zero slant to the line of sight, a gray color, and an impression of definite illumination. These reports are consistent with Metzger's, but suggest that the variable of phenomenal texture equality is not the same thing as his hypothetical microstructure. Our surface had a coarse physical texture; his plaster wall had a fine texture. It would seem, then, that the phenomenal quality of texture or pattern must be distinguished from the condition of the retinal image supposed to produce a surface.

When the illumination of the wall was gradually reduced, a level was reached at which all the properties mentioned were no longer reportable. The texture seemed to melt, the impression of hardness softened, the definite distance became foggy, and the impression of color was no longer separable from the impression of illumination. When the illumination was raised, these properties of a surface seemed to become definite again. They appeared to be in some degree linked together, although it is possible that different absolute thresholds for different properties might be determinable with a more elaborate setup.

Whether a continuous dimension from hardness to softness could have been discriminated in this situation is not possible to decide after the fact. A regular psychophysical experiment is needed.

The results are consistent with the hypothesis that the hardness of a surface is related to the steepness of the gradients of luminous intensity in the retinal image.

Effect of artificially focused and unfocused retinal images.—A large lighting fixture or globe, made of milk glass, was mounted in front of *O*'s eyes so as to fill the entire field of view. The shape of the outermost section was that of a flattened hemisphere. When moderately illuminated from outside, the visual field was homogeneous and no surface was visible. The translucent surface in front of the eyes, approximately 20 cm. distant, was sufficiently near flat so that a lantern slide could be projected on it from outside and the image would be in focus at the center but increasingly out of focus toward the periphery of the field. Photographic slides were prepared of various textured surfaces (a ploughed field, wallpaper, line patterns) and the projector was focused for the central region of this screen. In addition to the peripheral blur, the central image could also be gradually blurred by altering the projector lens.

When the projector beam was switched on, *O*'s impression of a luminous fog gave way to an impression of something a short distance in front of his eyes. Presumably the eyes accommodated and converged for the central portion of the photographic image. The percept could not, however, be called a surface comparable to the wall described above. Distance was definite, there was a certain hardness at the center of the field, and the quality of texture was reportable but no other surface properties were evident. The *O* did not mistake the image for a real surface, although this illusion is sometimes possible in experiments with an image on a flat translucent screen (2). The field became vaguer and perhaps softer toward the periphery as the gradients of intensity in the image became shallower.

When the center of the image was made to blur, the periphery became even more blurred. The impression of softness then increased over the whole field.

These results are consistent with the general idea that a sharply textured image makes for a surfacelike experience, but they also suggest that this formula is insufficient. The phenomenal texture observed was neither continuously extended, as in looking at a wall, nor bounded, as in looking at an object. Perhaps the alternatives of being either extended or bounded are part of the essential stimulus conditions for the impression of a visual surface.

Observations with a bounded region in the retinal image.—The texture-hypothesis in its crude form might lead one to predict that any bounded region of the visual field would be a surface when its image was differentiated or "speckled" and a film when its image was undifferentiated. The facts, however, contradict such an inference. An ordinary object on a background appears hard rather than filmy in everyday experience even when the structure of its surface is so homogeneous, or its distance is so great, that the corresponding retinal image must be homogeneous in effect. Likewise a rotating color disk has a homogeneous retinal image but yields the impression of a hard surface. Katz noted facts of this sort (4), and the contradiction can be shown experimentally.

A simple device for inducing a surrounding field of one kind of color and texture with a sharply bounded region of another color and texture is a large screen with an aperture or window in it. Under the name of a "reduction-screen" having an aperture of small visual angle, this setup is well known in the experiments on color perception.

Several cardboard aperture-screens were constructed with circular holes at the center varying in diameter from 1 cm. to 30 cm. The *O* fixated the aperture at a distance of 50 cm., usually with monocular vision and a motionless head. Behind the screen any of a number of different surfaces with different textures could be set up, at a further distance of 100–200 cm., in such a way as to fill the aperture.

Exploratory observations demonstrated that under some conditions, especially with the larger apertures, *O* saw a hard surface at some distance behind the screen, with a definite surface color and a definite texture. The perception was similar to that of looking out a window at, for instance, a surface of grass. Under other conditions, especially with the smaller apertures, *O* saw a soft penetrable color in the aperture itself, i.e., film color, or what has been called a "reduced color" (4). The phenomenon was not like that of a window but merely like a vague hole in a surface. Presumably in this event *O* was accommodating for the nearer rather than the farther surface. There were, however, more alternatives than these. Under some small-aperture conditions *O* saw a hard surface "on" the screen instead of a film "in" the aperture. This result seemed to be more likely when the aperture was much darker or much lighter than the screen, although the conditions for this, as for the other impressions, cannot be specified since the number of variables in the situation was large. In such cases the aperture

was sometimes reported to look like a disk of black or white paper pasted on the screen. It had no visible texture, but it nevertheless appeared to be a surface *with respect to hardness*.

This result seems to contradict the texture hypothesis. Evidently the hypothesis cannot be interpreted to mean that phenomenal hardness depends on retinal texture in the case of a sharply bounded region of the visual field. However, it is possible that the steep intensity gradient at the *contour* may have been the effective stimulus for the impression in this case rather than the texture within the contour. This interpretation becomes more likely if one remembers that a texture is itself composed of many small regions of the field and that it is the sharp margins or contours of these elements of texture which hypothetically produce the hardness of a continuous surface. It would not be surprising if the rule that held for a small region of the retinal image also held for a larger region of it. The suggestion is that the stimulus conditions for texture and for contour are fundamentally alike.

When the conditions of this setup were such as to arouse the "window" type of experience, a clear impression of edge was obtained. As might be expected, the jump in distance appeared more definite with binocular than with monocular vision, and when the head moved than when it was motionless. If fixation of a single eye with a motionless head were long continued, however, there seemed to be a tendency for the farther surface to come forward into alignment with the nearer surface, although retaining its surface qualities. The depth at the edge was then ambiguous. A few comparisons suggested that this result might be more likely when the texture of the two surfaces was different, e.g., when one was patterned and the other unpatterned. The situation obviously demands systematic investigation.

Effect of a variable aperture on the judgment of surface or film.—In order to investigate further the transition between the impression of surface and that of film, one of the variables in the aperture-screen setup was selected for more systematic study—the variable of aperture size. An iris diaphragm of the type used in cameras was mounted at the center of a large screen. Both the screen and the diaphragm were painted black. Since the aperture was continuously variable from a diameter of 2 mm. to about 50 mm. and the screen was set at a variable distance of 10, 20, or 30 cm. from *O*'s eye, a range of aperture sizes was obtainable from approximately 1 degree of visual angle to 24 degrees. The *O* sat with head fixed in a headrest, fixated the aperture, and reported its appearance. All vision was monocular. Five *O*s were used, all graduate students with training in this type of observation.

Behind the screen at a distance of something over 2 m. was set a textured surface of high reflectance. For two *O*s this consisted of cotton sheeting; for three others it was plasterboard with a coarser texture. In order to control accommodation, the eyes of all *O*s had been dosed with atropin (5% solution of homatropin) 40 min. before the experiment and the distance of the critical surface was adjusted for each *O* so that it was necessarily in focus. The effect of this procedure was to force accommodation for the surface and prevent accommodation for the aperture screen. It can probably be assumed that the transitions from surface to film were not accompanied by changes in accommodation, but were produced solely by a decrease in the size of the retinal image. The room was fully illuminated. At the larger apertures *O* simply saw a white textured surface through a circular hole in the black screen, the edges of the hole being blurred because of the far accommodation.

The intention of this experiment was to determine the minimum angular extent of a given texture which is required to see a surface. The aperture was either decreased from its maximum or increased from its minimum at any given distance until *O* reported a transition from surface to film or from film to surface. Ten ascending and ten descending judgments were obtained at each of the three distances from the screen. Each *O* was allowed to determine for himself the criterion for making his judgment, and each had been given a considerable degree of practice in making the distinction.

The results of the experiment were not quite as expected. Although the judgment was difficult to make, every *O* could do so and his angular threshold could be determined for the ascending and descending series and for the three different distances. Each individual *O* showed some consistency but the different *O*s gave widely different thresholds. The lowest threshold obtained for an *O* was in the neighborhood of 2 degrees and the highest in the neighborhood of 20 degrees. These differences were not related to the coarseness of the texture used. The suggestion was that each *O* had developed a different subjective standard of what he called a surfacelike impression. The only fact that emerged clearly was that the impression of a surface becomes more likely when the aperture is increased in size, and less likely when it is decreased. The implication is that this impression is not an entity but a variable, and that there is no clear point of difference between the impression of surface and that of film. It is possible that the accommodation response of the eye is a matter of all-or-nothing, but that

the retinal stimulation usually correlated with the response is a variable.

The aspect of stimulation being varied here was area, or number of adjacent visible elements of texture. Steepness of intensity gradients for these elements was held constant. The results indicate that the quality of hardness is a function of the *extent* of a differentiated image as well as its intensitive differentiation.

The contour of the aperture was always blurred in this experiment rather than sharp as it was in the previous setup, i.e., a shallow intensity gradient rather than a steep one. Accordingly, it is noteworthy that the film in the aperture was never reported to look like a hard surface on the screen as it sometimes was previously.

Discussion

Two conclusions, at least, are suggested by these exploratory observations, although many questions still remain in doubt. In the first place it is probably a mistake to assume that a phenomenal surface is an elementary impression of visual space. It is not a thing at all, but a variable of things. The hardness of a surface lies on a dimension having the quality of softness at the other extreme.

In the second place the visual hardness of a surface does not seem to be in psychophysical correspondence with texture as such. The texture-hypothesis as usually formulated is inadequate. A sharp contour in the retinal image seems to yield hardness whether it delimits a small speck of light or a large patch of light. The formula of the steepness of gradients of luminous intensity between regions of the image, however, gives promise of being valid.[4] The size of the regions referred to is unspecified, and is a matter for empirical study. The investigator must learn to think of the retinal image in terms of light rather than of objects which reflect light if he is to explain our perceptions of the latter. From this point of view the speckled image of a so-called texture and the contour image of a so-called figure differ only in that the former has many elements instead of one.

The problem of the stimulus conditions for an edge is similar to that for surface character except that more variables are involved. Perhaps edge is best conceived as a variable, not an entity. A true phenomenal edge bounds a phenomenal surface and it may be the case that one can never elicit the former without the latter.[5] Only a beginning has been made in the study of edge character during the present experiments, but they do suggest ways in which an investigator can go about making systematic experiments. There seem to be three basic ways of producing retinal images which arouse an edge. First is a setup which bisects the visual field, i.e., produces an unclosed edge in experience. Second is a setup which yields what might be called the *window phenomenon*, a closed edge with depth increasing from outside to inside the contour. Third is a setup which yields a closed edge with depth increasing from inside to outside

[4] Gradients of intensity are suggested in this statement as a first formulation although, in addition, gradients of wavelength might have to be considered. The ineffectiveness of a difference in hue without a difference in brightness to produce a visual contour (5), however, suggests that this complication may be minor. The *amount* of intensity difference as well as the gradient of this difference between adjacent regions may also have to be considered.

[5] The quality of *edge* is, in fact, a sort of upper threshold for the quality of slant. It is probably the impression which slant approaches as a limit when it increases from zero slant (a surface perpendicular to the line of sight) to an infinite slant (a surface parallel to the line of sight). As one fixates the flat face of an object, a cube for instance, and walks around it, the surface gets increasingly slanted (and the form gets increasingly foreshortened) until suddenly the surface disappears and becomes an edge. This relationship will be further discussed in a subsequent report.

the contour. This is allied to the classical *figure-ground phenomenon.* All these setups involve the impression of two planes of distance. The evidence suggests the hypothesis that under some circumstances an edge may degenerate into a mere margin or contour, in which case the two planes of distance become indistinct. The phenomenon of figure-on-ground has apparently been studied heretofore under these circumstances.

Surface character can be studied without the complicating effect of edges or of any boundaries save those of the visual field itself. The wall experiments demonstrate this possibility. The most interesting surfaces, however, are those exemplified by the flat face of an object and the ground behind the object (and various combinations of these). Experimentation with these requires the control of a large number of variables. In order to do so, it is wise for the experimenter first to spend some time taking an unprejudiced look at the visual world about him.

(Manuscript received September 24, 1951)

References

1. Gibson, J. J. *The perception of the visual world.* Boston: Houghton-Mufflin, 1950.
2. Gibson, J. J. The perception of visual surfaces. *Amer. J. Psychol.*, 1950, **63**, 367–384.
3. Gibson, J. J., & Waddell, D. Homogeneous retinal stimulation and visual perception. *Amer. J. Psychol.* (In press.)
4. Katz, D. *The world of colour.* London: Kegan Paul, Trench, Trubner and Co., 1935.
5. Liebmann, S. Über das Verhalten farbiger Formen bei Helligkeitsgleichheit von Figur und Grund. *Psychol. Forsch.*, 1927, **9**, 300–353.
6. Metzger, W. Optische Untersuchungen am Ganzfeld. II. Zur Phänomenologie des homogenen Ganzfelds. *Psychol. Forsch.*, 1930, **13**, 6–29.

PERSONAL AND ENVIRONMENTAL DETERMINANTS OF PERCEPTION IN A SIZE CONSTANCY EXPERIMENT [1]

JEROME L. SINGER [2]

University of Pennsylvania

In this investigation a fairly simple size constancy experiment was conceived of as a kind of model universe in which perceptual determinants heuristically classified in terms of the relative contribution of personal and environmental factors could be varied so as to permit study of their interaction in producing the ultimate perceptual response. Extending somewhat the view of Tolman and Brunswik (**15**), construction of a perceptual hypothesis may be considered as depending upon the relative clarity of environmental conditions and the organism's capacity and desire to delineate precisely the physical properties of objects. Translating this formulation to a size constancy procedure we can consider the distance of the standard stimulus from *S*, the angle of regard, or the stimulus content as environmental determinants. The demand made upon the perceiver for a more or less differentiated response through the medium of instructions represents the intermediate determinant. Here the situation, externally structured, depends upon the individual for interpretation and compliance. Finally, the perceiver's contribution, his capacities or habitual cognitive orientation, e.g., analytic or synthetic, represent the personal determinants. This series of experiments is designed to test hypotheses concerning the interaction of these personal, intermediate, and environmental determinants of perceptual response.

Experiment I

Problem.—In this preliminary experiment major emphasis was placed upon establishing the effect of primarily environmental determinants, the distance of standard stimuli from *S*, content of standard stimuli, and content of comparison stimuli. Thus, in a factorially-designed experiment, it could readily be predicted from previous research (**6, 14**) that increasing the distance of the standard stimuli from *S* would prove to be a significant source of variability. The importance of the distance determinant lies in the assumption made in Exp. III that increasing the distance between the standard stimulus and *S* decreases the relative structuredness of the field, thus enhancing operation of subjective factors in perception (**3, 17**).

Recently, there has been increased concern with the influence of the content of a stimulus upon its perceived size. Brunswik and his students (**6, 8, 10**) studied this problem, as did Ansbacher (**1**). A number of investigations (**3, 4, 12**) suggest the generalization that both meaningfulness and positive symbolic value of the stimulus content lead *S*s to overestimate stimulus size. In this experiment the use of a schematic "happy face" (Fig. 1) as one of four types of stimulus content should result in a significant influ-

[1] This paper is an abridged version of a dissertation offered in partial fulfillment of the requirements of the Ph.D. degree at the University of Pennsylvania under the supervision of Professors Francis W. Irwin, Miles Murphy, and James C. Diggory. The author is also indebted to members of the Psychology Department and to staff and trainees at the Philadelphia V. A. Clinic who cooperated in numerous ways.

[2] Now at Franklin D. Roosevelt V. A. Hospital, Montrose, N. Y.

ence of the content variable upon perceptual performance. When the standard is rather far from *S* the size-enhancing effect of meaningful content should partially counteract the gradual breakdown of constancy with increasing distance. Hence it is hypothesized that comparison stimuli equated by *S* to the distant standard should be larger when meaningful content appears on the standard than when nonsense figures or no content are employed.

Because of their proximity to *S* no effect of varying content of comparison stimuli may be expected. This variable has been included to rule out the necessity in subsequent research of employing content on the comparison squares identical with the content on the standards. Of greater import is the hypothesis anticipating a significant effect of individual differences in performance since this would suggest the operation of personal determinants. In addition some research by Klein, Meister, and Schlesinger (9) suggests that *Ss* may differ in the extent to which their perceptions are influenced by stimulus content. Hence, a significant interaction between *Ss* and stimulus content may emerge, some *Ss* proving more susceptible than others to the size-enhancement effect of meaningful stimulus content.

Procedure.—There were six *Ss*, four male graduate students and two male clerical workers, all ignorant of the purpose of the experiment. The *S* was seated in a corridor 35 × 15 ft. Directly to his front at distances of 10 and 30 ft. were two ringstands from which booklets of white squares, 5 × 5 in., were suspended. There were four squares in each booklet each with a different stimulus content, a blank white square, a square with a black-outlined circle of 3-in. diameter drawn on it, a poorly-articulated, meaningless figure constructed within a 3-in. diameter on a square, and a square with a schematic "happy face" constructed within a circle of 3-in. diameter (Fig. 1). The comparison stimuli consisted of four booklets, one for each type of standard stimulus content. Each booklet contained squares increasing in size in ¼-in. intervals from 2 to 6 in. The comparison stimuli were exposed on ringstands 5 ft. from *S* at an angle of 45° from the sagittal plane.

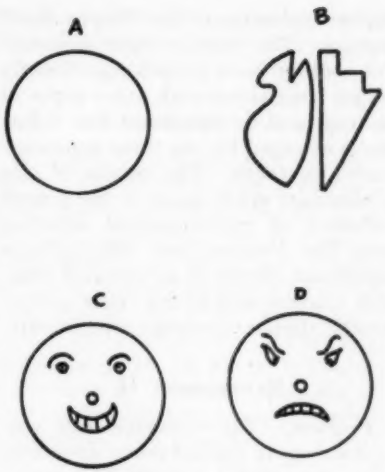

FIG. 1. Content of standard stimuli: A = Circle; B = Nonsense Figure; C = "Happy Face"; D = "Sad Face."

Instructions called upon *S* to choose a comparison square equal in size to a distant standard. The four standard stimuli were exposed twice in conjunction with each of the four types of comparison stimuli. Thus a B standard was presented twice at 10 ft. and twice at 30 ft. with each type of comparison stimulus content. Each *S* made 64 judgments in all with the order of presentation for each *S* randomized. The data were treated by analysis of variance using the data from replication as the estimate of error.

Results.—The results of Exp. I supported the main hypotheses. The source variables of subjects, distances, and content of standard stimuli yielded *F* ratios significant at the .01 level. Content of comparison stimuli had no significant effect upon size judgments. Of 11 possible interactions only two, *Ss* × standard stimulus content and *Ss* × distances, were significant at the .05 level. The effect of the standard content was due almost entirely to a pronounced tendency for *Ss* to give

higher estimates of the "happy face" square. The squares with meaningful content were judged significantly larger than those with other types of content and no significant size differences emerged for the three nonmeaningful contents. The results of this preliminary study point to the potent influence of environmental determiants like distance, but also indicate significant effects of meaningful stimulus content and of individual differences in the size constancy experiments.

Experiment II

Problem.—The objective here was to investigate the influence upon size perception of intermediate determinants, experimental instructions. Contrasting instructions were used to induce differing perceptual attitudes and to study the capacity of *Ss* to comply with such external demands under conditions of varied stimulus contents. The "Bet" instructions called for *S* to make every effort to estimate accurately the physical size of the distant standard. He could use knowledge, judgment, and rely fully on field cues. The "Look" instructions demanded that *S* impose a "functional reduction screen" on the field, ignoring field cues and avoiding the use of knowledge. On the basis of considerable earlier research (**7, 8, 10**) we may predict a significant influence of the instructions variable on perceptual performance with "Bet" instructions yielding values closer to constancy than "Look" instructions. Similarly, work by Klimpfinger (**10**) and Thouless (**14**) leads us to predict a significant interaction of *Ss* and instructions, some *Ss* showing greater ability or predisposition to comply with certain instructions than others.

The different stimulus contents in Exp. II included again the "happy face," nonsense figure, and blank square, with the addition of a "sad face." The latter content was included as a tentative check on the possibility that the size-enhancement effect of the "happy face" in Exp. I was a function of positive symbolic value rather than meaningfulness per se. Previous research (**4, 9, 12**) suggests that stimulus content should again prove to be a significant source of variation but that despite the predicted size-enhancement of the squares with meaningful content, no significant difference between the "happy face" and "sad face" would emerge.

There is increasing evidence that, when reliance on extensive objective cues is limited, *S* responds more readily to qualitative aspects of a stimulus (**5, 9**). In this investigation the objective attitude called for by the "Bet" instruction with its reliance on full field information should operate to counteract the size-enhancement phenomenon of meaningful stimulus content thus yielding a significant instructions × stimulus content interaction. Finally, to the extent that individual differences play a pervasive role in *S*'s compliance with instructions or in his response to meaningful stimulus content, a significant triple interaction, *Ss* × instructions × stimulus content, may be anticipated.

Procedure.—The design and procedure were identical with those of Exp. I with certain exceptions. The *Ss* were six female students who had only the slightest acquaintance with constancy experimentation. Standard stimuli were 4-in. squares with four types of content—"happy face," "sad face," nonsense figure, and blank (Fig. 1). The standard squares were exposed only at 25 ft. The comparison stimuli, blank squares increasing in units of ½ in. from 1½ to 5½ in., were exposed to *S*'s right at a distance of 5 ft. and at an angle of 30° from the sagittal plane. Each standard stimulus was exposed four times, twice in connection with each of the two types of instructions. Each *S* thus made 16 judgments in a sequence determined by prior randomization. The estimate of error for analysis of variance was once again based on the replication of *S*'s performance.

Instructions for the analytic, "Look" attitude were: "I want you to look at the white square to your front and then at a series of white squares exposed on your right. Tell me when one of the series looks to you to be equal to the one in front. Remember, I am interested in how they *look* to you and not when you judge them to be equal or use external comparisons to aid yourself. Try as much as possible to avoid being influenced by what you know of the various cue relationships and merely tell me when they look equal or the same."

Instructions for the objective, "Bet" attitude were: "I want you to look at the white square to

your front and then at the series exposed to your right. You are to compare the squares themselves for size. Now in doing so indicate when one of the squares to your right seems equal to the standard out there by using any methods of judgment or consideration in addition to their actual "look" or appearance. In other words, if you react as if you had a bet on these and wanted to be pretty certain they were both the same size, you'll be following these instructions."

Results.—Table 1 gives the mean size judgments for the 4-in. standard of the four types of stimulus content under the two types of instructions. The meaningful stimuli "happy face" and "sad face" yield higher values than the other types of content. The difference between the means of the meaningful and nonmeaningful content is significant at the .01 level. There is no significant difference between the two types of meaningful content. Analysis of variance reveals that the source variables for *Ss*, stimulus content, and instructions all yielded *F* ratios significant at the .01 level. The *Ss* × instructions and stimulus content × instructions variables were significant at the .01 level and the triple interaction of *Ss* × instructions × stimulus content was significant at the .05 level. All of these results support the hypotheses and only the *Ss* × stimulus content interaction failed to prove significant in this experiment as predicted. Again, the size-enhancing influence of meaningful stimulus content emerges, tempered, however, by the extremely potent influence of the perceptual set induced by the instructions.

TABLE 1

MEAN SIZE JUDGMENTS OF STANDARD STIMULI IN EXP. II

Stimulus Content	Instructions	
	"Look"	"Bet"
Blank square	3.08	3.74
Nonsense figure	3.08	3.75
"Happy face"	3.35	3.77
"Sad face"	3.30	3.67

EXPERIMENT III

Problem.—The next phase of investigation involved the role of two personal determinants, experimentally-induced frustration, representing a momentary emotional disturbance, and thinking introversion-extroversion, representing a long-term personality characteristic. Previous literature (**7, 10, 14, 16**) on individual differences in constancy had delineated two habitual modes of perceptual orientation, the introverted, subjective, or analytic and the extroverted, objective, or synthetic. The cognitive introvert, perhaps because of excessive immersion in self and analysis of perceptions and thought processes, is likely to show less tendency toward constancy. The extreme cognitive extrovert, on the other hand, tends to be more dependent on the stimulus characteristics of the external environment and is likely to show perceptual responses approaching constancy despite instructions of the "Look" variety. We can predict a significant effect of the introversion-extroversion variable in Exp. III. If two distances are employed, however, we may expect an interaction of introversion-extroversion and distance, the personality factor emerging more clearly at the farther distance. This is in keeping with the general hypothesis stated earlier, that personal needs or predilections are more potent as the perceptual task increases in difficulty or ambiguity.

The employment of experimentally-induced frustration is an extension to the constancy situation of previous work (**2, 11, 13**) which has suggested that frustration leads to more primitive, undifferentiated modes of behavior or perception. In this experiment, frustrated *Ss* should

prove less able to comply with the analytic, "Look" instructions, which require a decided effort to separate out the standard square from the background, avoiding reliance on field cues. Frustrated *Ss* should respond more globally and in keeping with habitual tendencies to maintain constancy. A significant effect of the frustration source variable should be anticipated, frustrated *Ss* showing higher values than controls. A differential effect of distance may also be expected, the perceptual influence of frustration proving more effective with the standard further away.

Procedure.—The *Ss* were 36 male undergraduates, the 18 extreme thinking introverts and 18 extreme thinking extroverts of 68 respondents to Guilford's Inventory of Factors STDRC. The mean difference in scores on the Thinking Introversion Scale for the two groups was 27.9 points within a range for 68 respondents of 45 points. The groups did not differ significantly in age. The *Ss* were randomly assigned to four groups: frustrated introverts, control introverts, frustrated extroverts, control extroverts. The *Ss* underwent the following sequence of procedures: (*a*) two control perceptual trials with standard square at 35-ft. distance; (*b*) intervening task, frustration or sorting (for controls); (*c*) experimental perceptual trials, two at each distance.

Control *Ss* performed an innocuous card-sorting procedure for 15 min. Frustrated *Ss* were given an unsolvable modification of the Vigotsky Concept-Formation Test. They were informed that they were being compared with norms based on performance of psychotics. The *E* expressed concern, surprise, and anxiety over their failure to complete the task, readily communicating this mood to *Ss* who showed numerous clinical signs of emotional disturbance. After 15 min., *Ss* were immediately led outside for the experiment trials. A subsequent interview to obtain introspections and to allay anxiety was conducted.

The constancy procedure employed 8-cm. blank white squares which were exposed at distances of 10 and 35 ft. Comparison squares ranged from 2 to 10.25 cm. in .25-cm. intervals. Instructions were of the "Look" type (Exp. II). Separate analyses of variance were carried out for each distance and for the control perceptual trials.

Results.—In contrast with controls, frustrated *Ss* manifested numerous signs of disturbance during perceptual trials and reported tendencies toward impulsive perceptual reactions. The mean size judgments in centimeters for each condition are presented in Table 2. The data support the hypotheses, with extroverts showing larger values (significant at the .01 level) prior to frustration while differences between control introverts and extroverts at both 10 and 35 ft. are also significant at the .01 level. Analysis of variance yields an *F* ratio of 22.96 (significant at the .01 level) for introversion-extroversion at 35 ft. prior to frustration, an *F* of 7.73 (significant at the .01 level) following frustration at 35 ft., and an *F* of 4.97 (significant at the .05 level) at 10 ft. The controls and frustrated *Ss* do not differ significantly prior to induction of frustration with analysis of variance of pretrials at 35 ft. yielding an insignificant *F* ratio. Following frustration, analysis of variance yields no significant *F* ratio at 10 ft. although the difference between control and frustrated introverts barely misses significance at the .05 level. For the 35 ft. distance the frustration source variable yields an *F* of 26.07 (significant at the .01 level). The results are thus in accord with the hypotheses, frustrated *Ss* showing higher values than controls at the farther distance. The control *Ss* improved somewhat in carrying out the instructions while frustrated *Ss* were less successful in complying with instructions even to the extent of the supra-constancy manifested by frustrated extroverts. With the greater ambiguity of cues at

TABLE 2

MEAN SIZE JUDGMENTS OF STANDARD STIMULI IN EXP. III

Group	Control	Experimental	
	35 ft.	10 ft.	35 ft.
Control introverts	6.85	7.25	6.52
Frustrated introverts	6.81	7.72	7.86
Control extroverts	7.77	7.88	7.29
Frustrated extroverts	8.00	7.87	8.48

35 ft. the frustrated extroverts, who prior to frustration already showed strong tendencies toward constancy, overcompensated for distance.

Experiment IV

Problem.—The final experiment investigated the joint operation of the personal, environmental, and intermediate perceptual determinants, frustration, stimulus content, and instructions. With frustrated and control *Ss* making size judgments of squares with meaningful and blank content in accordance with "Look" or "Bet" instructions, significant effects of the source variables should again be anticipated. Similarly, the stimulus content × instructions interaction should prove significant as it was in Exp. II. If the relationship of instructions to the frustration variable is considered, another interaction is likely to result. The marked effect of change in instructions, representing an intermediate determinant, should prove even more potent than frustration, representing a personal determinant. Hence the hypothesis was that changing instructions from "Look" to "Bet" should have a greater effect upon control *Ss* than upon frustrated *Ss*, particularly since the potency of environmental factors imposes an upper limit even on impulsive guessing by frustrated *Ss* during "Bet" instructions. An interaction between frustration and instructions could, therefore, be predicted, the difference in size judgments between frustrated and control *Ss* decreasing sharply as instructions change from "Look" to "Bet."

Experiment II provided evidence that the optimal size-enhancement effect of meaningful stimulus content occurs under analytic, "Look" instructions. Since in Exp. III frustrated *Ss* proved less capable of complying with "Look" instructions, a double interaction, frustration × stimulus content, may be predicted. Control *Ss* should prove more susceptible to the influence of meaningful content than frustrated *Ss*. This differential effect of stimulus content should prove to be more marked when instructions are taken into account. Control *Ss* may be expected to show larger values in response to meaningful content than do frustrated *Ss* when instructions are of the "Look" variety. No such differences in the effect of stimulus content on controls and frustrated *Ss* should emerge when "Bet" instructions are employed. A significant triple interaction, frustration × stimulus content × instructions is, therefore, to be anticipated.

Procedure.—*Ss* were 56 male undergraduates without knowledge of constancy experimentation. They were randomly assigned to eight experimental groups of seven *Ss* each. Each group underwent the constancy procedure under a different combination of the three variables, frustration or control with meaningful or blank content following "Look" or "Bet" instructions.

Experiment IV was almost identical in design with Exp. III. Standard stimuli were two 9-cm. white squares, one blank and one with the "happy face." Squares were exposed 30 ft. from *S* so that content was visible. Blank comparison squares ranged from 3 to 10.5 cm. in intervals of .25 cm. Prefrustration trials were conducted using a blank square and "Look" instructions. The intervening task and frustration were exactly as in Exp. III and the "Look" and "Bet" instructions as in Exp. II.

Results.—Analysis of variance of pre-experimental trials for the eight groups shows no significant influence of the original groupings. Table 3 presents means in centimeters for *Ss* in all experimental conditions. Results of the analysis of variance of experimental trials are presented in Table 4.

The data in Table 4 reveal that all source variables and interactions, with the exception of frustration × stimulus content, were significant as predicted while inspection of Table 3

TABLE 3

MEAN SIZE JUDGMENTS OF STANDARD STIMULI IN EXP. IV

Condition	Stimulus Content	Instructions	Mean Distance	
			Pre	Post
Control	Blank	"Look"	7.47	7.25
Control	Meaningful	"Look"	7.38	8.18
Frustrated	Blank	"Look"	7.39	8.52
Frustrated	Meaningful	"Look"	7.50	8.90
Control	Blank	"Bet"	8.00	9.06
Control	Meaningful	"Bet"	7.52	9.01
Frustrated	Blank	"Bet"	7.68	9.32
Frustrated	Meaningful	"Bet"	7.22	9.47

TABLE 4

ANALYSIS OF VARIANCE OF POST-FRUSTRATION SIZE JUDGMENTS IN EXP. IV

Source	*df*	Variance	*F*
Frustration	1	20.6429	29. **
Instructions	1	64.2858	89. **
Stimulus content	1	4.5715	7.02*
Frustration × instructions	1	3.4999	5.32*
Frustration × content	1	1.4107	2.17
Instructions × content	1	3.0178	4.63*
Frustration × content × instructions	1	3.5008	5.37*
Error	48	.6514	

* Significant at the .05 level.
** Significant at the .01 level.

indicates that the general direction of differences was in accordance with the hypotheses.

DISCUSSION

Although the specific results of these experiments involving small *N*s and simple procedures, must be evaluated with caution, the series in general seems to point up the usefulness of heuristic models in perceptual-personality research. Certainly, the dimensions of stimulus content need much more detailed exploration than was feasible in this series to delineate better the specific basis for the apparent tendency of *S*s to overestimate the size of a white square because a schematic face appears on it. Similarly, precise delineation of the role of emotional disturbance in perceptual response is needed to determine whether its effect is merely to prevent efficient compliance with instructions or to influence *S*s towards excessive reliance upon field cues.

In general, the data lend further support to the important principle that personality characteristics show their clearest influences upon perceptual response in direct relation to the increased difficulty or ambiguity of both the environmental situation and the performance demanded of the perceiver.

SUMMARY AND CONCLUSIONS

Employing a simple size constancy procedure, four experiments were conducted to observe the interaction of personal and environmental determinants of perceptual response. The *S*s made size judgments of distant squares under conditions which included changes in the distance and in meaningfulness of stimulus content, two types of experimental instructions, and following emotional disturbance.

1. Experiment I presented evidence to support the hypothesis that meaningful stimulus content (a schematic "happy face") led to a significant enhancement in perceived size of a standard square compared with squares containing a nonsense figure, a circle, or blank surface.

2. Experiment II provided results suggesting that this effect of meaningful stimulus content was operative when analytic rather than synthetic or objective perceptual attitudes are demanded of *S*s. The *S*s also differed markedly in susceptibility to the influence of meaningful content and in mode of compliance with instructions.

3. In Exp. III, *S*s chosen as extreme thinking introverts and extroverts differed significantly in perceptual response, extroverts showing values closer to constancy. Similarly, frustrated *S*s showed values closer to constancy than controls. The effect of these personal determinants was greater when the standard was farther away from *S*.

4. In Exp. IV, the effect of frustration was shown to be more striking when analytic rather than objective perceptual attitudes were called for. Similarly, frustrated *Ss* proved less susceptible than controls to the effect of meaningful stimulus content when analytic perceptual attitudes were required.

In general, the results lend some support to the view that personal factors play a significant role in perception, particularly as the performance expected of *S* increases in difficulty or ambiguity.

(Manuscript received July 30, 1951)

References

1. Ansbacher, H. Perception of number as affected by the monetary value of objects. *Arch. Psychol., N. Y.*, 1937, No. 215.
2. Barker, R., Dembo, T., & Lewin, K. Frustration and regression. *Univer. Iowa Stud. Child Welfare*, 1941, **18**, No. 1.
3. Bruner, J., & Goodman, C. Value and need as organizing factors in perception. *J. abnorm. soc. Psychol.*, 1947, **42**, 33–44.
4. Bruner, J., & Postman, L. Symbolic value as an organizing factor in perception. *J. soc. Psychol.*, 1948, **27**, 203–208.
5. Bruner, J., Postman, L., & Rodriguez, J. Expectation and judgment of color. Paper read at the Eastern Psychol. Assn., Worcester, 1950.
6. Brunswik, E. Die Zugänglichkeit von Gegenständen für die Wahrnehmung und deren quantitativ Bestimmung. *Arch. ges. Psychol.*, 1933, **88**, 357–418.
7. Henneman, R. A photometric study of the perception of object color. *Arch. Psychol., N. Y.*, 1935, No. 177.
8. Holaday, B. Die grössen Konstanz der Sehdinge bei Variationen der inneren und ausseren Wahrnehmungsbedingungen. *Arch. ges. Psychol.*, 1933, **88**, 419–486.
9. Klein, G., Meister, S., & Schlesinger, H. The effect of personal values in perception: an experimental critique. Paper read at Amer. Psychol. Assn., Denver, 1950.
10. Klimpfinger, S. Über den Einfluss von intentionaler Einstellung und Ubung auf die Gestaltkonstanz. *Arch. f. d. ges. Psychol.*, 1933, **88**, 552–598.
11. Korchin, S., Singer, J., & Ballard, R. The influence of frustration on the reproduction of visually-perceived forms. *Personality*, 1951, **1**, 54–66.
12. Lambert, W., Solomon, R., & Watson, P. Reinforcement and extinction as factors in size estimation. *J. exp. Psychol.*, 1949, **39**, 637–641.
13. Postman, L., & Bruner, J. Perception under stress. *Psychol. Rev.*, 1948, **55**, 314–324.
14. Thouless, R. Individual differences in phenomenal regression. *Brit. J. Psychol.*, 1932, **22**, 216–241.
15. Tolman, E., & Brunswik, E. The organism and the causal texture of the environment. *Psychol. Rev.*, 1935, **42**, 43–77.
16. Weber, C. The relation of personality trends to degrees of visual constancy correction for size and form. *J. appl. Psychol.*, 1939, **23**, 703–708.
17. Witkin, H. The nature and importance of individual differences in perception. *J. Pers.*, 1949, **18**, 145–170.

SOME RELATIONS BETWEEN STIMULUS PATTERNS AND PERFORMANCE IN A CONTINUOUS DUAL-PURSUIT TASK [1]

PAUL M. FITTS AND CHARLES W. SIMON

The Ohio State University

In many skilled tasks requiring the adjustment of controls in response to stimuli provided by visual displays, the relations between the controls and the displays are not directly observable. The *S*s must discover a set of rules or hypothetical relations, or at least establish a set of responses that are compatible with the particular control-display relations that exist. In tasks where *S* must discover these relations, performance is considerably influenced both by the direction of movement and the spatial relations between the particular displays and controls employed. Results from studies of the ability of *S*s to learn to utilize some of the many possible control-display arrangements have been the subject of recent reviews by Mitchell and Vince (**14**) and Fitts (**7**, pp. 1306–1311).

The following variables have been found to influence performance in this type of task: (*a*) whether the display and its control move in the same or different planes of space (**1, 16**); (*b*) whether the display is above or below, to the right or to the left, in front of or behind its associated control (**1, 15, 16, 17**); (*c*) whether or not, when its associated control is moved, the direction of the movement of a display is the "expected" one, i.e., conforms to a population stereotype (**1, 9, 10, 11, 13, 14, 17**). It also appears that performance is more markedly affected by such factors when the task is complex and requires frequent change of set (**14**).

The investigations reported here were concerned with a limited topic within the area outlined above. Three experiments were undertaken to determine the interrelation of stimulus-pattern variables in a moderately complex pursuit task. The variables studied were (*a*) the direction of alignment of instrument pointers, (*b*) vertical versus horizontal separation of instruments, and (*c*) the distance between instruments.

Apparatus and Task

The Grether Dual-Pursuit Apparatus (**12**) was used in all three studies (see Fig. 1). The *S*s watched the pointers on two different instrument dials and tried to keep them within designated limits by making appropriate adjustments of two rotary control knobs, one of which was operated by each hand.

In this apparatus a motor-driven cam provides problem inputs to two differential gears. The movements of *S*'s control knobs provide additional inputs. The differentials act as subtractors, continuously computing the differences between the problem inputs and *S*'s corrective movements. The outputs from the differentials, which are equivalent to error, are transmitted electrically to two instrument dials, where they appear as pointer deflections; they also are transmitted mechanically to wiper arms that move across metal contacts, and activate scoring clocks. The instrument dials are marked so that *S*s can see when the pointers are "on target."

Work and rest periods are governed by a sequence timer, which also triggers a warning

[1] These experiments were carried out at Antioch College under a contract supervised by the Psychology Branch, Aero Medical Laboratory, USAF Air Materiel Command. At that time the first author was on the staff of the Aero Medical Laboratory and the second author was on the staff of Antioch College. The authors collaborated in planning the studies and preparing this report; Mr. Simon supervised the collection and analysis of the data.

FIG. 1. The Grether Dual-Pursuit Apparatus. The control box and the two control knobs are shown at *A;* the instrument panel, cut away to reveal the method of mounting the instruments, is shown at *B;* the *E*, control unit, and scoring clocks are shown at *C*.

sound before the beginning of each trial. A scoring clock cumulates time whenever both pointers are simultaneously "on target." This clock reading was used as the criterion of proficiency throughout the present investigations.

The instrument dials used in all experiments were 3 in. in diameter, black with standard white pointers, and mounted on a black instrument panel. Four lights, set in reflectors, one at each corner of the panel, provided an even, non-glare source giving an apparent brightness of 30 ft.-lamberts. The *S*'s eyes were 28 in. away from, and at the same height as, the mid-point between the two instruments.

The panel on which the control knobs were mounted was tilted at a 45° angle away from *S* (see Fig. 1), so that the knobs could be grasped comfortably. Adjustable armrests were used. One control knob was above and to the right of the other, the line formed by the two knobs making an angle of 45° with the base of the panel. The upper right-hand knob controlled either the upper or the right-hand instrument (depending on whether the two displays were separated vertically or horizontally), and the lower left-hand knob controlled either the lower or the left-hand instrument. This arrangement of controls was used throughout the three experiments. The *S*s found this arrangement an expected one, and easily associated each control with its related instrument.

EXPERIMENT I: THE RELATIVE EFFICIENCY OF TEN DIFFERENT POINTER–POSITION PATTERNS

Procedure.—Ten stimulus configurations were compared in the first experiment. These patterns are shown schematically in Fig. 2. Two directions of dial separation, vertical (V) and horizontal (H), and five different positions of the two pointers, were studied.[2] The pointers were aligned, respectively, at 3, 6, 9, and 12 o'clock, and with their tips counterpoised (directed toward each other). These will be referred to hereafter as the 3, 6, 9, 12, and X positions. In Exp. I the tips of the two pointers were separated by a distance of 8 in.

Twenty male and 20 female college students served as *S*s. All had normal visual acuity. None had operated the apparatus before. They were given the following instructions while seated at the apparatus:

"These pointers will move back and forth. Try to keep them within the white reference marks by moving these knobs in the appropriate directions. Try them. (At this point *S*s were allowed to turn the knobs and to note the resulting movements of the pointers.)

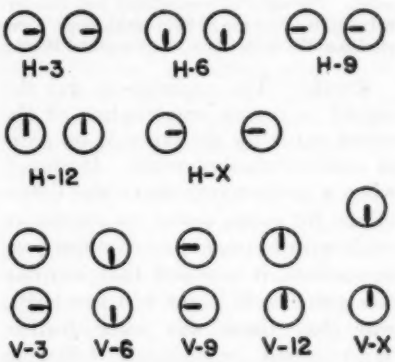

FIG. 2. The ten different stimulus patterns employed in Exp. I

[2] Reference throughout the paper to "pointer position" will be used for convenience to refer to the position of the target or zero point in the quadrant of the dial in which the pointer is functioning. This was at one of the four cardinal positions. In reality, the position of the pointer was constantly shifting and only rarely, except through the efforts of *S*, was it positioned exactly at any cardinal point.

"Try to keep both pointers on their respective white marks *at the same time*. The time you keep them *both* on will be recorded. You will have ten trials, each followed by a short rest period. After the fifth trial you will be given a longer rest period. Do not turn the knobs during the rest period. A buzzer will warn you a second before the rest period is over. Experience suggests that it is more efficient to move your eyes back and forth between the two dials than to attempt to watch both dials out of the corners of your eyes. So move your eyes as you work.[3]

"The position of the pointers will be changed from trial to trial."

One group always worked with the dials separated horizontally. The other group worked with them separated vertically. Hereafter these will be called the "H" and the "V" groups. The *Ss* were assigned to the two groups at random, with the restriction that each group have an equal number of men and women.

Each *S* was given ten 3-min. trials. The first 15 sec. of each trial were treated as a warm-up period, and only the last 165 sec. were scored. The *Ss* rested for 5 min. between the fifth and sixth trials; after all other trials a 90-sec. rest period was used.

A different pointer-position pattern was presented on each of the first five trials. These five patterns were then presented again in reverse order. The order of presentation was counterbalanced between the 20 *Ss* in each group. Completion of the series of ten trials required 50 min.

Results.—The experiment was designed to permit combination of the scores made by the same *Ss* on pairs of counterbalanced trials. However, when a performance curve was determined for mean scores on successive trials without regard to stimulus configurations, it revealed that learning was quite rapid in the first five trials, and that there was some further improvement immediately following the mid-session rest period. The last five trials, however, showed a relatively stable performance level. Since the error term would have been seriously inflated by combining initial- and final-trial data, only the data for the last five trials were analyzed.[4]

TABLE 1

MEAN PER CENT OF TIME SIMULTANEOUSLY "ON TARGET"

(N = 20 *Ss* in Each Group)

Pointer Position	Group H		Group V		Mean for Both Groups
	Mean	*SD*	Mean	*SD*	
3 o'clock	42.5	7.5	33.8	7.7	38.1
6 o'clock	39.7	7.1	36.3	8.9	38.0
9 o'clock	45.7	6.3	35.5	7.7	40.6
12 o'clock	43.5	5.9	39.2	8.4	41.4
Counterpoised	45.8	5.8	38.4	8.4	42.1
Mean for all positions	43.4		36.6		40.0

Mean performance scores and standard deviations based on the final five trials for each condition are given in Table 1. The results of an analysis of variance for these data (6, pp. 284–302) are summarized in Table 2. The variance between mean scores for horizontal and vertical instrument alignment was significant ($p < .01$) when tested by between-subject variance for *Ss* with the same dial separation. For the task employed in this experiment it can be concluded that performance with horizontally-separated dials is superior to performance with vertically-separated dials.

Using the pooled interaction of *Ss* and pointer positions as the estimate of error variance for correlated groups, the variance due to the interaction between instrument alignment and pointer position was found to be significant ($p < .01$). This indicates that performance with a particular pointer alignment is in part dependent upon

[3] The instruction to move the eyes was given in order to secure uniformity in the way different *Ss* worked; actually the *Es* have no data on whether or not *Ss* do better when they are instructed to move their eyes in this task.

[4] In subsequent experiments the combining of counterbalanced trials in this task was made possible by using shorter test trials that permitted frequent variation in testing conditions, and by continuing the experiment over a period of several days.

instrument alignment. The converse is also true.

Performance with the counterpoised pointers was not significantly different from that with the best of the other alignments for Groups H and V considered separately and ranked first when data for both groups were combined.

For Group H performance with the 9 o'clock pointer position was significantly better than with the 12 o'clock alignment ($p < .05$) and performance with the 3 o'clock position was significantly better than with the 6 o'clock alignment ($p < .02$). For Group V the converse was true. The 12 o'clock pointer alignment was superior to the 9 o'clock ($p < .01$), and the 6 o'clock alignment was significantly better than the 3 o'clock ($p < .05$).

For both the horizontal and the vertical instrument alignments considered separately, 9 o'clock pointer position was superior to the 3 o'clock, and the 12 o'clock was superior to the 6 o'clock position. Three of these differences were significant at the .02 level, but one (9–3 o'clock, vertical) was not significant ($p < .15$).

Since the data were not suitable for a three-dimensional analysis of variance, a subtraction procedure was adopted as a means of testing the effect of pointer position independent of instrument alignment. The two counterpoised conditions were not used in this analysis. Mean scores were obtained for the four possible combinations of (*a*) horizontal or vertical instrument alignment with (*b*) corresponding or noncorresponding pointer alignment. Subtracting the adjusted means from the appropriate original scores for individual *Ss* removed the effects of instrument alignment and of corresponding versus noncorresponding pointer alignment. Subjecting the residual scores to a simple analysis of variance for correlated groups gave a value for *F* of 7.57 which with 3 and 117 *df* is significant at better than the .01 level.

TABLE 2

ANALYSIS OF VARIANCE IN EXP. I

Source of Variation	*df*	Mean Square	*F*
Between dial separations	1	22.99	9.82
Between *Ss* with the same dial separation	38	2.34	
Between pointer positions	4	1.39	*
Interaction of pointer positions and dial separation	4	.87	8.13
Interaction of pooled *Ss* and pointer positions	152	.107	

* See the text for a discussion of a further analysis of this factor.

TABLE 3

VALUES OF *t* FOR ADJUSTED MEAN DIFFERENCES BETWEEN POINTER POSITIONS IN EXP. I*
($df = 39$)

	6 o'clock	9 o'clock	12 o'clock
3 o'clock	−.517	2.794**	3.349**
6 o'clock		3.311**	3.865**
9 o'clock			.555

* Plus values indicate that the mean performance of the pointer position on the abscissa was greater; minus values, that it was smaller.
** $p = .01$ or better.

The results of this analysis of pointer positions, with other effects removed, are given in Table 3. It is apparent that the 9 and 12 o'clock positions are superior to the 3 and 6 o'clock positions, but that the members of each pair do not differ significantly from each other.

Intercorrelations between scores made during the 165-sec. trials at the five pointer positions were computed. The median intercorrelation for Group H was .74, and for Group V was .86. These intercorrelations indicate that factors common to an individual's ability to perform on the pursuit task contribute a good deal of weight to his

score no matter what stimulus pattern is used.

Experiment II: Extended Learning Study Comparing 9 and 12 o'clock Pointer Positions

A second experiment was undertaken to determine more precisely the relative effectiveness of the 9 o'clock vs. the 12 o'clock pointer alignments when used with horizontal and vertical instrument alignments. In addition, the experiment was designed to determine whether or not the differences attributable to pointer position would persist throughout an extended learning period.

Procedure.—The effective width of the "on-target" scoring area was reduced from that employed in Exp. I in order to provide a task in which improvement would continue throughout a period of extended practice. The distance between instruments was held constant at 8 in.

One group (H-V), consisting of four men and four women, practiced for 15 days with the instruments aligned horizontally and for 5 additional days with them separated vertically. Another group (V-H), consisting of three men and three women, practiced for 15 days with instruments aligned vertically and for another 5 days with them separated horizontally. All *S*s practiced alternately at 9 and 12 o'clock conditions.

Trials lasted for 70 sec. The first 10 sec. constituted a warm-up period during which no score was recorded, but this was not known to *S*. Rest periods of 50 sec. were used. Twenty trials were given each day. The sequence of conditions was counterbalanced between *S*s and between days. The instructions to *S*s were equivalent to those in Exp. I.

The *S*s were told their scores at the end of each trial. Midway through each daily series of trials they were complimented on their progress and encouraged to try to do even better. Each day they were also shown their learning curves. It appeared that *S*s were highly motivated throughout the four weeks of the experiment.

Results.—Learning curves for the extended practice sessions are shown in Fig. 3. The results agree with the previous findings in indicating the superiority of the vertical-vertical and the horizontal-horizontal alignment patterns. Means of Group H-V favored the 9 o'clock position for each of the first 15 days of practice with horizontal instrument alignment, while the means of Group V-H favored the 12 o'clock position on each of the first 15 days.

The data for each *S* during the initial 15 days were averaged over successive blocks of 5-day periods, giving three sets of data for each *S*. The differences between 9 and 12 o'clock positions for individual *S*s were then tested. All 18 comparisons for Group V-H favored the 12 o'clock condition and all were significant at the .01 level or better. Twenty-three out of 24 comparisons for Group H-V favored the 9 o'clock position and 18 of these were significant at the .05 level or better.

The smallest absolute and the smallest relative differences due to stimulus

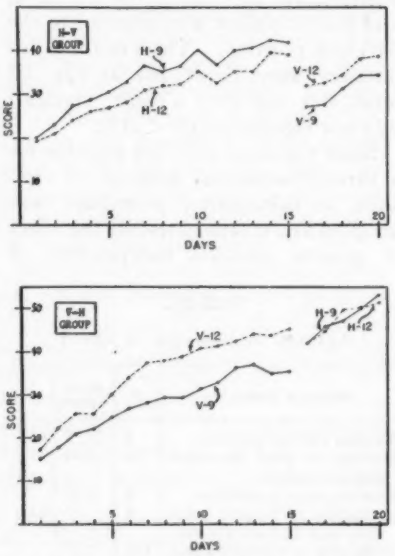

Fig. 3. Learning curves in Exp. II. Each point on each curve is the average of all trials for a given day (10 trials per *S*).

pattern factors occurred on the first of the 15 days for both groups. The level of significance of the differences increased from the first to the second 5-day practice period and showed no appreciable change during the third 5-day period.

When instrument alignments were reversed during the fourth 5-day period, all 14 *Ss* evidenced a shift in performance in a direction consistent with the results of Exp. I. All the evidence indicates, therefore, that the effects of stimulus-pattern factors persist throughout an extended training period.

Experiment III: the Effect of the Extent of Instrument Separation on Performance

The third experiment was conducted to determine how performance is affected by the distance between displays. Results were secured for both horizontal and vertical instrument alignment.

Procedure.—The same apparatus was used as in the previous experiments. New panels were constructed to permit separations of 4, 8, 12, or 16 in. between the centers of the instruments.

Twelve college men, with normal vision, were used as *Ss*. None had ever operated the apparatus before. The instructions were comparable to those used in Exp. I and II. In order to encourage different *Ss* to use uniform procedures, all were told to move their eyes back and forth between the two instruments.

Each *S* was tested at the same time of day for six consecutive days. During these daily sessions, *S* worked with both vertical and horizontal alignments and at all four distances. Two 130-sec. trials were given for each of the eight conditions (a total of 16 trials per *S* per day). The sequence of test conditions was counterbalanced within and between *Ss* and between days.

The 12 o'clock pointer alignment was used when instruments were separated vertically and the 9 o'clock alignment was used when they were separated horizontally, since results of the two preceding experiments showed that these were optimum conditions. The "on-target" tolerance limits were the same as those in Exp. I. Unknown to the *Ss* no score was recorded during the first 10 sec. of each 130-sec. trial. The *Ss* rested for 1 min. between trials; after every fourth trial there was a 2-min. rest period.

TABLE 4

Performance Scores for Four Distances of Horizontal and Vertical Instrument Separation in Exp. III

($N = 12$)

Separation of Instruments (Inches)	Horizontal		Vertical	
	Mean	*SD*	Mean	*SD*
4	61.2	6.4	51.8	7.2
8	48.3	6.0	42.0	5.6
12	43.7	5.4	36.2	5.8
16	39.2	5.3	33.1	5.6

Results.—In all respects performance was better when the dials were closer together than when they were farther apart. Group results are summarized in Table 4. For both alignments, scores became successively worse with each 4-in. increment in the distance between instruments. Successive differences, based on group data, were all significant.[5] In addition, the score of each of the 12 *Ss*, without exception, decreased with each increment in the distance between instruments along both the horizontal and the vertical axes.

In all cases performance was significantly better for a horizontal separation of a given amount than for a vertical separation of the same amount. This relation held, without exception, for the average scores of each individual considered separately, as well as for the means of the groups. The absolute superiority of the horizontal

[5] The standard error of the mean difference for these conditions was computed to be 1.034 by using the variance of the pooled interactions of *Ss* × pointer positions as the best variance estimate for all groups. This estimate of standard error was used to compute the *t* values, in place of an estimate of standard error of the difference computed separately for each pair of conditions.

over the vertical condition increased as the distance between instruments increased.

Individual differences within the 12 *Ss* remained relatively stable with variations in alignment and instrument separation. Correlations between scores for vertical vs. horizontal alignment at equal distances ranged from .88 to .95; correlations between adjacent distances along the same axis varied from .92 to .98.

Discussion

The three experiments reported here were primarily empirical in nature, and the chief significance of the data results from their indication of systematic relations between stimulus pattern variables and performance in a visual-motor task. The results have wide implications for a common class of problems in applied experimental psychology, in which the psychologist is interested in discovering optimum relations within and between displays and their related controls for the class of tasks in which *Ss* cannot observe these relations directly but must discover them.

The data have further significance in indicating the degree to which certain sensory and perceptual data can be generalized to the prediction of performance in pursuit tasks. It appears, on the one hand, that the effects of the distance and the direction of instrument separation are most directly accounted for by visual receptor characteristics. On the other hand, the effects of variations in pointer position, and the interaction between instrument and pointer alignment, appear to indicate the importance of more complex pattern-perception processes.

Instrument distance and instrument alignment effects indicate that peripheral vision is of considerable importance in perceptual-motor tasks. As the separation of stimulus objects along the vertical and the horizontal axes of the visual field was increased, a gradient of decreasing performance was obtained which resembles the gradients reported for threshold and visual acuity functions (**3**). Phenomena relating to the visual musculature and to the mechanics of eye movements are probably of less importance in this type of task than is peripheral vision. Since the duration of eye fixations in this dual-pursuit task are of the order of .5 sec., as compared with durations of the order of .02 to .04 sec. for saccadic movements in this amplitude range (**5**), the actual time taken by the latter would be expected to have a relatively small direct effect on performance.

The interaction between instrument and pointer alignment suggests that performance in a multi-stimulus task is facilitated when the over-all stimulus pattern is as simple, or unified, as possible. In these experiments performance was better when the instrument and pointer alignments corresponded, i.e., lay along a single continuous line. Under nonalignment conditions the pointers lay along two parallel lines, thus forming a less unified total stimulus pattern.

The experimental finding of most significance for a general theory of perceptual-motor space is probably that regarding the superiority of the left and upper sectors of a circular display. This superiority, which is independent of alignment factors, does not appear to be predictable from any established principles of space or pattern perception. However, several investigators have reported similar results in widely different experiments.

In addition to the present experiments, which utilized a continuous pursuit task, performance has been found to be superior in an instrument check-reading task when pointers are aligned in the 9–12 o'clock sectors of a dial (**2, 18**). Eye fixations in checking instrument panels also have been reported to be more frequent on the upper-left areas of the visual field (**19**). Dallenbach (**4**) found greater relative subjective brightness with lights located either to the left or above a fixation point.

The present writers have proposed that the superiority of the 9–12 o'clock alignment is consistent with the fact that display-control relations may be perceived in one of two ways—as movements in curvilinear or rectilinear space (**8**). Assume that a point on a display or a control moves in an arc of relatively short length and long radius. This movement may be perceived either as curvilinear (clockwise–counterclockwise) or as rectilinear (right–left, up–down). For example, when we rotate the steering wheel of an automobile clockwise, we usually verbalize this action by stating that we are turning the wheel to the right, and people understand what we mean because they assume that we are referring the movement to a point at the top of the steering wheel. If the movement of a display is perceived as rotary it seems reasonable that it should also be perceived as calling for a rotary response, or if perceived as rectilinear (up–down, right–left), as calling for a translatory movement response. The significant point, for the present results, is that whenever these alternative conceptual schemes lead to identical interpretations and response predictions for any given sector of a circular display, then they must lead to contradictory predictions when pointers are aligned in the opposite sector, i.e., when the pointers have rotated 180°.

A complete analysis of the conditions leading to control-display ambiguities in pursuit tasks requires consideration of display and of control patterns and their interactions. A complete formulation of these conditions is beyond the scope of the present paper. The present data, however, are consistent with the limited hypothesis that sectors of incompatability should be opposite to (180° away from) the sectors of best performance. The superiority of the left and upper sectors may be peculiar to the particular set of control-display relations employed here, although the agreement of our findings with several related experiments (**2, 4, 18, 19**) suggests a more general significance of this finding.

Subsequent experiments in the present series (**15**), in which a lever was used in place of a rotary control, have shown that the spatial orientations of the control (such as the orientation of a lever), as well as the spatial patterns of the stimulus, have a significant influence on performance, and that there is a strong interaction between control and display orientation. The available data suggest, therefore, that stimulus-pattern effects, response-pattern effects, and interactions between and within these two sets of variables, all influence the efficiency of performance in perceptual-motor tasks.

Summary

Various visual stimulus patterns formed by different arrangements of instruments and pointers were studied by means of a continuous, dual-pursuit task in which *Ss* adjusted two control knobs in order to keep two pointers within designated limits.

In these experiments fixed control arrangements were used. The controls were below their associated displays, rotated in the same planes of space as the instrument pointers, and a clockwise rotation of a control resulted in a clockwise rotation of its related pointer. These are common or "expected" relations.

The results from three experiments are in agreement in indicating that *Ss* give significantly superior performance when instruments are (*a*) close together, and (*b*) aligned horizontally. When instrument alignment factors are eliminated, pointer alignment in the 9 and 12 o'clock sectors of a dial gives significantly better performance than does alignment at 3 and 6 o'clock. However, if instrument alignment is permitted to interact with pointer alignment, performance is superior when pointers are aligned at 9 o'clock for horizontally-separated instruments, at 12 o'clock for vertically-separated instruments, or when the pointers are counterpoised.

The results are of importance for a common problem in applied experimental psychology, the optimum spatial arrangement of displays. The demonstration that differences in initial performance, due to pointer position, actually increased slightly during 15 daily 1-hr. practice sessions has general significance. Pointer configuration effects apparently cannot be explained on the basis of transient unfamiliarity with a given stimulus pattern, but must be attributed to well-established differences in the perceptual-motor capacities for responding to different kinds of stimulus patterns.

An hypothesis, based on the principle of simplicity in the total stimulus field, is proposed to account for the superiority of pointer alignment over nonalignment. Another hypothesis, based on a distinction between the interpretation of movement in reference to curvilinear or rectilinear space, is proposed to account for the superiority of performance when pointers are aligned in the left and upper sectors of a circular display.

(Manuscript received August 3, 1951)

References

1. Carter, L. F., & Murray, N. L. A study of the most effective relationships between selected control and indicator movements. In P. M. Fitts (Ed.), *Psychological research on equipment design*. Washington: U. S. Government Printing Office, 1947.
2. Connell, S. C., & Grether, W. F. Psy-

chological factors in check reading single instruments. USAF Air Materiel Command Memorandum, Dayton, Ohio, Report No. MCREXD-694-17A, 1948.

3. Crozier, W. J., & Holway, A. H. Theory and measurement of visual mechanisms. II. Threshold stimulus intensity and retinal position. *J. gen. Physiol.*, 1939, **22**, 351–364.
4. Dallenbach, K. M. Position versus intensity as a determinant of clearness. *Amer. J. Psychol.*, 1923, **34**, 282–286.
5. Dodge, R., & Cline, T. S. The angle velocity of eye movements. *Psychol. Rev.*, 1901, **8**, 145–157.
6. Edwards, A. S. *Experimental design in psychological research.* New York: Rinehart, 1950.
7. Fitts, P. M. Engineering psychology and equipment design. In S. S. Stevens (Ed.), *Handbook of experimental psychology.* New York: Wiley, 1951.
8. Fitts, P. M., & Simon, C. S. The arrangement of instruments, the distance between instruments, and the position of instrument pointers as determinants of performance in an eye-hand coordination task. Wright Air Development Center, Dayton, Ohio, AF Technical Report 5832, 1952.
9. Fitzwater, J. T. A study of the effects of rest pauses in perceptual-motor learning involving compensatory pursuit. Unpublished M.A. thesis, Ohio State Univer., 1948.
10. Gardner, J. F. Direction of pointer motion in relation to movement of flight controls. USAF Air Materiel Command, Dayton, Ohio, Technical Report No. 6016, 1950.
11. Grether, W. F. Efficiency of several types of control movements in the performance of a simple compensatory pursuit task. In P. M. Fitts (Ed.), *Psychological research on equipment design.* Washington: U. S. Government Printing Office, 1947.
12. Grether, W. F. A dual compensatory pursuit apparatus for use in psychological research. USAF Air Materiel Command, Dayton, Ohio. Technical Report No. 6036, 1950.
13. Loucks, R. B. An experimental evaluation of the interpretability of various types of aircraft attitude indicators. In P. M. Fitts (Ed.), *Psychological research on equipment design.* Washington: U. S. Government Printing Office, 1947.
14. Mitchell, M. J. H., & Vince, M. A. The direction of movement of machine controls. *Quart. J. exp. Psychol.*, 1951, **3**, 24–35.
15. Simon, C. W. Instrument-control configurations affecting performance in a compensatory pursuit task. I. Interaction between type and location of the control and the position of the instrument pointer. Wright Air Development Center, Dayton, Ohio. A. F. Technical Report No. 6015, 1952.
16. Warrick, M. J. Direction of movement in use of control knobs to position visual indicators. In P. M. Fitts (Ed.), *Psychological research on equipment design.* Washington: U. S. Government Printing Office, 1947.
17. Warrick, M. J. Effects of motion relationships on speed of positioning visual indicators by rotary control knobs. USAF Air Materiel Command, Dayton, Ohio. Technical Report No. 5821, 1949.
18. Warrick, M. J., & Grether, W. F. The effect of pointer alignment on check-reading of engine instrument panels. USAF Air Materiel Command, Dayton, Ohio, Memorandum Report No. NCREXD-694-17, 1948.
19. White, W. J. The effect of dial diameter on ocular movements, speed, and accuracy in check-reading groups of simulated engine instruments. USAF Air Materiel Command, Dayton, Ohio. Technical Report No. 5826, 1949.

THE ROLE OF STIMULUS MEANING (*m*) IN SERIAL VERBAL LEARNING[1]

CLYDE E. NOBLE[2]

State University of Iowa

One of the earliest experimental tests of the doctrine that "meaningfulness facilitates learning" was carried out in 1914 by Lyon (17). Using nonsense syllables, digits, prose, and poetry, and by recording the time required to master equal amounts of these materials, he found that learning time decreased in the order given. Subsequent investigations (19) involving a variety of procedures corroborated Lyon's findings, apparently verifying "common sense" expectation. These pioneering studies, however, shared the fundamental defect that no one had specified pre-experimentally a quantitative index of "meaningfulness" which was independent of the anticipated performance data.

In 1928 Glaze (6) laid the foundation for a rigorous test of the classical proposition when he worked out his "association value" for nonsense syllables. This concept was defined as the proportion of *S*s who respond positively to a given item within a 3-sec. period, indicating one or more free associations. Later, Hull (12), Kreuger (15), and Witmer (33) developed similar procedures.

In an unpublished thesis in 1929, Melton (21) reported an inverse relationship between Glaze association values and difficulty in serial learning. His *S*s learned 20-item lists at a 2-sec. rate by the spelling-anticipation method to a mastery criterion of one perfect trial. The three lists had mean association values of 0%, 46.7%, and 100%. The mean trials to mastery for the first cycle, with the absolute and relative variabilities shown in parentheses, were 40.75 (SD = 19.9; V = 48.8), 27.41 (SD = 10.44; V = 38.1), and 19.91 (SD = 7.81; V = 39.2), respectively. Close examination of Melton's mean values reveals a slight curvilinearity in what we may call the *difficulty-meaning relationship*.

Of the four association value scales cited above, the only one of known reliability is that of Hull: r = .64 (N = 20). The possibility exists, therefore, that when difficulty is plotted as a function of association value, the comparatively low precision of the abscissa scale may obstruct the discovery of the exact functional relationship. There is the further fact that, except for chance letter combinations, this scaling technique has not been extended to actual English words. Certain consequences of this obviously restricted range of meaningfulness have necessitated the assumption that nonsense syllables and 3-letter words lie on the same continuum (2, 18, 22, 26); but it is at least doubtful that the conventional methods possess adequate sensitivity in the upper levels to be useful. What is required is a scale of high reliability sampling an extensive range of stimuli

[1] This paper is part of a dissertation submitted to the faculty of the Department of Psychology of the State University of Iowa in partial fulfillment of the requirements for the Ph.D. degree, June 1951. The author is indebted to Prof. Kenneth W. Spence for his advice and criticism. A portion of this paper was read before the American Psychological Association, September 1951.

[2] Now at the Perceptual and Motor Skills Research Laboratory, Human Resources Research Center, Lackland Air Force Base, San Antonio, Texas.

by a method which evaluates the number of associations elicited by each item.

To facilitate the determination of the difficulty-meaning relationship is the principal reason for undertaking the present investigation. It represents a sequel to a previous study in which the author (25) developed a rational scale of meaningfulness (*m*), whose properties lend themselves to more precise quantitative treatment. Briefly, *m* was defined as the mean number of continued written associations given to a particular word within a minute. The scale consisted of 96 dissyllabic nouns and paralogs, was standardized on 119 cases, and had an intergroup reliability coefficient of .975.

Concerning the second objective, there is much research which indicates that distribution of practice facilitates the attainment of mastery (1, 9, 19, 29, 34). However, it is not clear whether this variable is related to stimulus meaning. Similarly, certain studies on the Ward-Hovland effect (7, 8, 30) have indicated that short-term reminiscence occurs under these conditions, although there is evidence to the contrary (11, 27, 28). But again, no data are available which systematically relate recall phenomena to stimulus meaning. Four investigations (2, 23, 24, 31), in fact, have reported absences of reminiscence when actual English words were used rather than nonsense syllables. Regarding serial position effects, the facts are equally obscure. Reductions in center position errors following rest are generally found for all materials (2, 7, 8, 14, 23, 24), but the relationships with meaning have not been worked out.

The quantification of meaning which led to the *m*-scale thus permits an examination of the relationships among meaningfulness, so defined, and learning difficulty, work-distribution, reminiscence, and intraserial interference. A knowledge of these relationships should clarify the role of stimulus meaning in serial verbal learning.

Method

The present investigation followed a commonly-employed technique with respect to variables known to be relevant to this class of learning phenomena (1, 3, 4, 19, 29, 32, 34).

The typical procedure involves: (*a*) 12 items, (*b*) 2-sec. interitem intervals, (*c*) 6-sec. intertrial intervals, (*d*) Yale instructions (7), (*e*) pronunciation-anticipation method with correction for errors, and (*f*) the filling of a 2-min. rest interval with color-naming at a 2-sec. rate.

The controls involved constructing the lists so that: (*g*) the mean number of units (letters) was approximately equal, (*h*) serial position effects were equalized, and (*i*) modified Müller-Schumann rules were satisfied (34). The variables of pre-experimental learning ability and prior practice were controlled by means of a treatments × levels design. These controls will be discussed more fully in the following sections.

Apparatus.—The apparatus consisted of a Missouri-type memory drum, glazed white cloth tape, a color sequence of 12 different hues applied to the tape, three practice lists of 12 three-syllable adjectives, and three experimental lists of 12 two-syllable nouns taken from the extremes and central portion of the *m*-scale (25). The memory drum was mounted in a screen, with *E* and *S* on opposite sides. The exposure window of the drum permitted six verbal and two color lists to be mounted in tandem, thus facilitating changes from one treatment to another during the experiment. The hues were arranged in a different random sequence for each list. The practice lists were of formally different appearance from the dissyllables comprising the experimental lists in order to minimize any effects of proactive inhibition on the latter task. Two of these practice lists were constructed from words high in the Thorndike-Lorge frequency tables, as used earlier by Noble (24), while one list was made up of paralogs invented by *E*. This was intended to give *S*s practice not only in learning lists but also in learning to pronounce unfamiliar verbal symbols.

The three experimental lists were constructed in accordance with the following rules: (*a*) no words with identical initial or final syllables within a single list, (*b*) no homophones, synonyms, antonyms, assonants, or alliterates within a list, (*c*) no initial syllables identical with the

final syllable of the preceding item, and (*d*) no adjacent items constituting a well-known phrase. In addition, each item appeared once in every ordinal position for each list; i.e., list order was counterbalanced, but sequence remained invariant.

As a further control over intralist heterogeneity, the provisional lists were submitted to a group of five judges (graduate students specializing in verbal learning) for critical appraisal. Their examination removed a number of unsatisfactory sequences. The mean list *m*-values ($\bar{m}$) for the low, medium, and high $\bar{m}$ lists were 1.28, 4.42, and 7.85, respectively. The mean number of letters per list was 5.6, 6.0, and 6.0, in that order.

Finally, in order to test for the stability of the experimental variable, a *familiarity* rating schedule using the same 96 items as in the *m*-scale was administered to a sample of 28 undergraduate men and women at the State University of Iowa. This device yielded simple ordinal scale values which were covariant with *m*. When the data for the 36 items in the three experimental lists were extracted from this sample (testing trait) and compared with the appropriate values on the *m*-scale (criterion trait), the rank-difference correlation coefficient was .95. This evidence indicated that the *m*-scale, though developed on a military population, could be used quite rigorously with the population represented in the present investigation.

Procedure.—The learning method was a standard serial anticipation procedure, with correction for errors. The following scoring criteria were adopted (**24**): a response was scored incorrect (*a*) if it was preceded by two successive identical anticipatory errors, or (*b*) if the item was mispronounced or only partially enunciated, or (*c*) if an otherwise correct response coincided with the appearance of the response word in the aperture of the memory drum.

The *S*s were 72 male and female students at the State University of Iowa, enlisted on a volunteer basis. There were 52 men and 20 women. Their ages ranged from 19 to 54, with a mean age of 25.2 years. All were naive with respect to serial learning and were not informed of the purpose of the investigation. They were given standard instructions and were requested neither to discuss the experiment nor to rehearse the practice lists between sessions. The general procedure was to administer the three practice lists in counterbalanced order and sequence on Day 1, and to administer an experimental list approximately 24 hr. later on Day 2.

The three practice lists were learned to a criterion of 7/12 correct, one of the two non-paralog lists including a 2-min. rest pause of color-naming upon reaching the criterion. Following the rest period, *S*s resumed learning until the 7/12 criterion was again reached. This regimen was intended to provide practice in shifting to and from the color-naming procedure. The interval between lists was 5 min., during which time *E* engaged *S* in conversation. It may be noted that the practice lists were counterbalanced as a control for any unequal sequence or order effects on the subsequent day's learning. The performance data from Day 1 also served as a control or leveling variable, on the basis of which *S*s were classified into three ability levels. These levels (*L*), defined by the total trials required to reach 7/12 mastery for the three practice lists, were as follows: Slow (>35), Average (26–35), and Fast (<26).

The present investigation utilized a double classification factorial design for the analyses of variance. To study the difficulty-meaning relationship a 3 × 3 treatments × levels design was set up, in which *S*s were randomly assigned to treatments ($\bar{m}$) within ability levels of the control variable (*L*). Since *S*s were further assigned to list orders within each treatment at random, the order effects were confounded with individual differences. Twenty-four *S*s were assigned to each treatment, two to each order. The distributed practice- and reminiscence-meaning relationships, in turn, were analyzed by 3 × 2 factorial designs, in which the rest-nonrest categories were analogous to the level classifications in the former design. In this case, 12 *S*s were already randomly assigned to each cell or treatment combination.

Results

Performance as a function of $\bar{m}$.—Figure 1 shows the performance curves for the three principal conditions of the experiment. Successive correct anticipations are plotted as functions of mean trials for the three $\bar{m}$ parameters, where $\bar{m}$ denotes the mean *m*-value of a 12-item list. The curves for the corresponding rest groups are seen to reach the 12/12 criterion slightly ahead of the nonrest groups.

Figure 2 exhibits the difficulty-meaning relationship for successive criteria of correctly anticipated items. These performance level criteria are denoted by the parameters accompanying the curves.

A more precise conception of this latter relationship is afforded by graph-

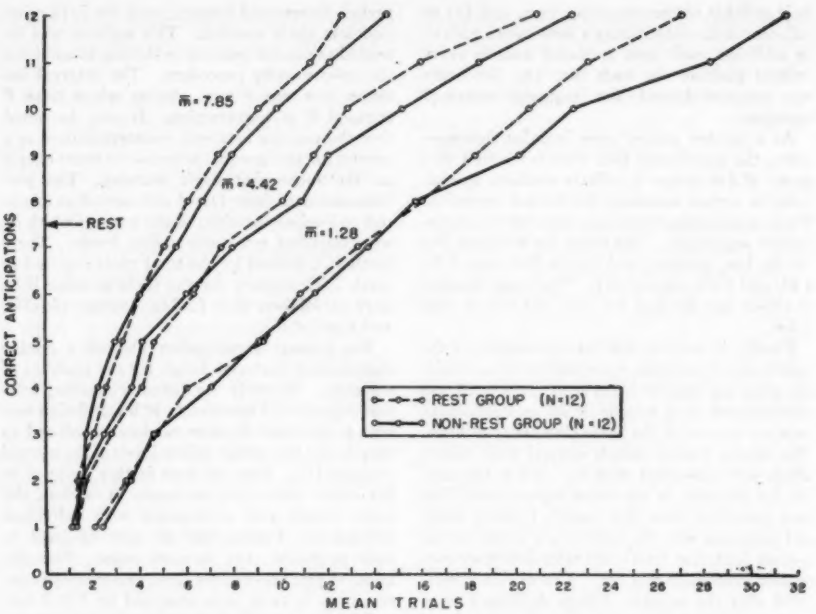

FIG. 1. Mean number of trials required to anticipate correctly successive numbers of items with and without a 2-min. rest pause of color-naming introduced after reaching the criterion of 7/12 correct. The $\bar{m}$ parameters denote average meaningfulness of 12-item lists.

ing the cell means from the treatments × levels analysis. In Fig. 3 are shown the difficulty curves to a 7/12 criterion as functions of $\bar{m}$ for the three pre-experimental ability levels (L). There is an apparent interaction between treatments and levels; i.e., between $\bar{m}$ and L.

Exact tests of the significance of the $\bar{m} \times L$ interaction and of the main effects of treatments averaged over levels are vitiated by the fact of heterogeneous variance. Using Bartlett's test (16), a χ^2 of 70.8 was obtained which, for 8 df, is highly significant ($p < .001$) of unequal variances within cells. This finding was to be expected in view of Melton's (21) work cited earlier, and others have confirmed the inverse relationship between variability and meaningfulness.

Considering that the established relationships operate at cross purposes with the present experimental design, one may evaluate the difficulty-meaning functions by Lindquist's (16) approximation rule. This rule asserts that if the ratio $\sigma^2_s/\sigma^2_{R \cdot C} > 4R$, where R and C denote rows and columns respectively, then the rank order of the treatment effects is identical for all levels. The present ratio has a value of 13.6, which is greater than 12. The usefulness of this test, which assumes interaction, is conveyed by Fig. 3. The apparent trend may not validly be tested with reference to the F distribution, but it seems reasonable to adopt a nonmetric test of significance for this purpose. Accordingly, the Mann-Whitney (20) U-test was applied to the distributions for the low-$\bar{m}$ and high-$\bar{m}$ lists; those having $\bar{m}$-values of 1.28 and 7.85 respectively. This procedure, which tests the null hypothesis that the cumulative frequency distributions being compared are equivalent, indicated a significance ratio of 5.11, which is highly significant ($p < .001$). Since the means for the 7/12 criterion shown in Fig. 3 consistently decrease with increasing $\bar{m}$-value, the observed trend may be regarded as statistically significant.

Distribution of practice and $\bar{m}$.—Inspection of Fig. 1 has suggested that the introduction of a 2-min. rest pause

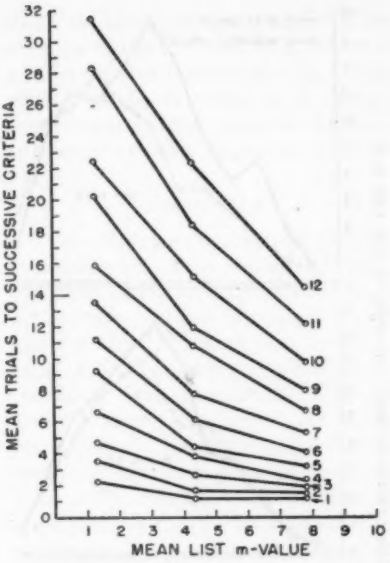

FIG. 2. The difficulty-meaning relationship as a function of successive criteria. Each point on the curves from 1/12 to 7/12 represents the performance of 24 *Ss*, while those from 8/12 to 12/12 represent the performance of 12 *Ss*.

of color-naming upon reaching a criterion of 7 items correct facilitates the eventual attainment of 12/12 correct. A 3 × 2 factorial design was set up for these data, where the columns represented *m*-value as before and the rows the rest and nonrest conditions.

TABLE 1

ANALYSIS OF VARIANCE OF TRIALS BETWEEN CRITERIA OF 7/12 AND 12/12 CORRECT WITH AND WITHOUT A 2-MIN. REST PAUSE AT 7/12 CORRECT

Source of Variance	*df*	Sum of Squares	*F*
Columns (*m*)	2	735.085	10.89**
Rows (rest-nonrest)	1	78.125	2.32*
Interaction $R \times C$	2	30.582	<1.00
Within cells	66	2227.083	
Total	71	3070.875	

* $p > .100$
** $p < .001$

TABLE 2

ANALYSIS OF VARIANCE OF ITEMS RECALLED ON FIRST TRIAL FOLLOWING CRITERION OF 7/12 CORRECT WITH AND WITHOUT 2-MIN. REST PAUSE AT 7/12 CORRECT

Source of Variance	*df*	Sum of Squares	*F*
Columns (*m*)	2	18.700	<1.00
Rows (rest-nonrest)	1	13.350	<1.00
Interaction $R \times C$	2	0.523	<1.00
Within Cells	66	168.413	
Total	71	200.986	

The analysis of variance is presented in Table 1. The test of interaction between list meaningfulness and practice distribution yielded an $F < 1.00$, which is not significant. The test of treatment differences by columns (*m*-value) resulted in an F of 10.89, which is highly significant ($p < .001$), but the observed difference between the row means (rest vs. nonrest), to the contrary, was not significant. The obtained F of 2.32 is associated with a probability $.10 < p < .20$. These data,

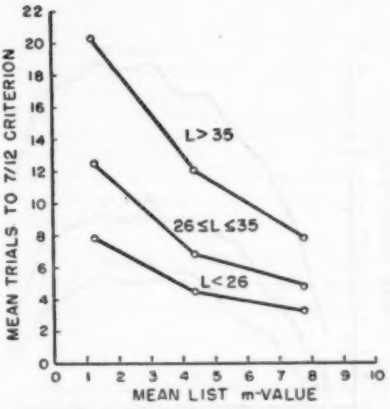

FIG. 3. The difficulty-meaning relationship as a function of pre-experimental ability level (*L*). Each point represents the performance of 8 *Ss*. The parameters denote total trials to 7/12 for three practice lists on preceding day.

representing a higher mastery level than did those of the previous analyses, exhibited no apparent problem of heterogeneous variance.

Reminiscence and $\bar{m}$.—To investigate the relationship between stimulus meaning and optimal conditions for reminscence, a factorial design similar to the preceding one was adopted. In this case the column and row treatments were identical, but the responses being analyzed were numbers of items recalled immediately versus those recalled 2 min. after the trial on which a 7/12 criterion was first reached. Table 2 presents the analysis of variance. The test of interaction resulted in an F less than 1.00, which is not significant. The tests of differences in the column means ($\bar{m}$-value) and in the row means (rest vs. nonrest) were likewise not significant; both F ratios were less than unity.

Serial position effects and $\bar{m}$.—Figure 4 exhibits the mean errors in learning to a 7/12 criterion as a function of serial position for the three experimental values of $\bar{m}$. There is a perfect inverse relationship between errors and meaningfulness.

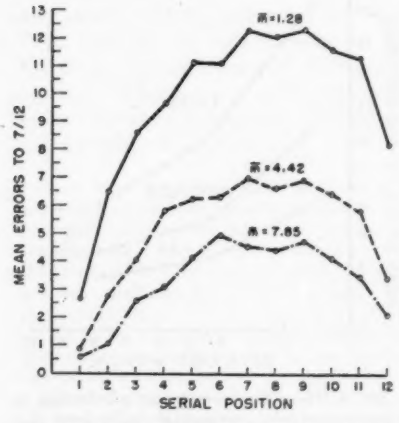

FIG. 4. Serial position effects as functions of $\bar{m}$-value. Each curve represents the performance of 24 *Ss*.

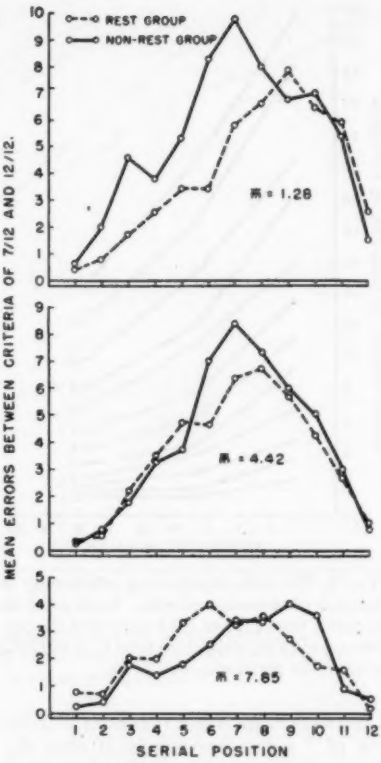

FIG. 5. Serial position effects as functions of rest for various $\bar{m}$-values. Each curve represents the performance of 12 *Ss*.

To evaluate the effects of rest on serial errors, it is customary to plot the error frequencies for the control and experimental groups following attainment of the criterion at which the rest condition is introduced. Such a comparison is shown in Fig. 5. The serial position errors incurred with and without rest between the criteria of 7/12 and 12/12 are graphed for the three values of $\bar{m}$. It will be noted that the error-savings are greatest for the list of lowest $\bar{m}$-value. Attention is in-

vited to the interesting shift in the maxima of the rest condition curves as a function of stimulus meaning. This effect may be described as a change, following rest, from negative skewness to symmetry with increasing $\bar{m}$-value.

Discussion

During learning to mastery, the variable $\bar{m}$ has been shown to have a highly significant effect upon performance, as indicated by the non-metric trend analysis for the 7/12 data and by the F ratio of 10.89 derived from Table 1 for the 12/12 data ($p < .001$). The evidence presented in Fig. 2 and 3 suggests that the fundamental difficulty-meaning relationship is of the form $Y = ae^{-bX}$; where $a \equiv$ a parameter denoting either the learning criterion or ability level, $b \equiv$ a parameter denoting the slope, and where Y and X are variables denoting difficulty and list meaningfulness respectively. Stated in other words, difficulty in serial learning is a decreasing exponential function of list m-value.

Since the curves in Fig. 3, if extrapolated, all have positive, real Y-intercepts and are presumably asymptotic to the X-axis, an exponential equation seems a rational one with which to describe them. Considering the ordinate, no 12-item list is expected to be impossible to learn, nor are any *Ss* likely to demonstrate a zero ability level. As for the abscissa, the curves should have slopes of approximately zero by an m_∞-value of 25.7, the theoretical limit of the m-scale imposed by the maximum speed of writing.[8]

[8] The value of m at infinity was estimated by reference to data obtained from a sample of undergraduates at the State University of Iowa by Dysinger (5), who found that the mean rate of successively writing the clause "Now is the time for all good men . . ." was 55 words per min. ($N = 161$; $r_{12} = .83$). Since Dysinger's *Ss* averaged 3.12 letters per word, and those of Noble 6.30 in their responses defining the m-scale, appropriate calculations for $m_\infty = (55) \cdot (3.12)/6.30 = 25.65$ words per min.

In the absence of additional points, of course, a rigorous test must be left to subsequent research. The present hypothesis is only heuristic.

Owing to the complications inherent in the unequal effects of the ordinal positions of items in a serial list, this hypothesis cannot be evaluated clearly with the present learning method. For obvious reasons a paired-associate technique more profitably might be used to determine whether the hypothesis holds for *specific* S–R connections. In addition to concomitant variation, one might vary the m-values of the items considered as stimuli and as responses independently. This procedure would provide information as to whether it is primarily the m-value of a stimulus or of a response—or, perhaps, of an interaction between the two—which differentially affects performance.

An unpublished dissertation by Sheffield (26) is a case in point. In a paired-associate situation he used nonsense syllables and three-letter words to define two qualitative attributes of "nonsense" (N) and "meaningful" (M) items respectively. Performance curves were increasingly facilitated as a function of these categories in the stimulus and response positions in the order: $N \rightarrow N$, $M \rightarrow N$, $N \rightarrow M$, $M \rightarrow M$. Sheffield's results are thus consistent with an S–R interaction hypothesis about the difficulty-meaning relationship. With respect to an interference theory, his data also revealed the greatest absolute and relative errors for the lists containing the largest proportion of nonsense syllables. This relationship held also for extralist intrusions and was especially marked for lists having nonsense items in the response positions.

The adoption of a quantitative index of meaning in a study like Sheffield's would permit more exact and complete tests of the present hypothesis regarding the difficulty function. A reaction threshold curve could also be determined. An extensive paired-associate investigation using m-values is urgently needed.

The use of a treatments × levels design was intended to increase the precision of the exponential hypothesis formulated above. Because of the limitations imposed upon the F test of interaction by heterogeneous variance, the observed interaction of $\bar{m}$ and L in Fig. 3 can only be considered

graphically. There is an apparent inverse relationship between pre-experimental learning ability (L) and sensitivity to differences in stimulus meaning. A tentative hypothesis may be proposed that slow learners are more sensitive to differences in meaningfulness than are relatively faster learners.

This investigation has clearly shown that varying $\bar{m}$-value bears no necessary relationship to the introduction of a 2-min. rest, so far as either savings in resuming learning to mastery *or* reminiscence is concerned. Furthermore, there is no significant interaction between $\bar{m}$ and these two response variables. With respect to the matter of the 2-sec. presentation rate, this variable seems to be importantly related to reminiscence, at least when nonsense syllables are used. Shipley (27) found that a 2.3-sec. rate was not associated with reminiscence, and Hovland (8) obtained nonsignificant decrements in recall with a 4-sec. rate. Concerning the effects of rest upon subsequent trials to mastery, the reports are that rest is generally facilitatory (19, 29, 34). Hovland (7, 8) found significant work-savings, but with adjectives, Melton and Stone (23) and Noble (24) obtained differences which were not significant, although in the expected direction. The present results, therefore, confirm those of all previous studies which have not used nonsense syllables.

It can be seen that these findings afford no support for the author's (24) hypothesis that m is inversely correlated with reminiscence. Considering the available evidence, one may adopt an alternative hypothesis that reminiscence is unique to the learning of nonsense *syllables;* i.e., three-unit items consisting of a vowel flanked by two consonants. But how shall one account for the recent Hovland-Kurtz (11) finding of a decrement under standard Ward-Hovland conditions? Another alternative would be that the phenomenon is associated only with highly practiced and highly motivated *Ss*. Hovland's *Ss* would certainly satisfy these criteria, since they were extensively counterbalanced and paid for their services. Hovland himself noted that "during the early stages of learning the color-naming has a disruptive effect on recall, but with further practice this effect disappears" (7, p. 220). Still other hypotheses might be entertained, such as subtle differences in instructions among various *Es*, lack of proportionality among m and association-value indices, or—in despair—the ubiquitous error variance. Whatever tack one chooses, it is at least unequivocal that reminiscence has occurred to date *only* with nonsense syllable material.

Of greater theoretical interest are the serial position effects observed in this investigation. From a postulated inverse relationship between associative inhibition, I,[4] and $\bar{m}$-value, a positive relationship between $\bar{m}$ and rate of learning was derived by Noble (24). The habit interference notion was based primarily upon the studies of Lyon (17) and McGeoch (18). The author did not then know of the unpublished papers of Melton (21) and of Sheffield (26). In view of their data and particularly the present findings of both absolute and relative intraserial interference effects as inverse functions of $\bar{m}$-value, Noble's 1950 pos-

[4] The construct I has some properties in common with what Hull *et al.* (14) termed inhibitory potential in that I has a subtractive effect upon ${}_{S}E_{R}$; but I is not to be confused with Hull's (13) I_R, *reactive inhibition*. Perhaps the symbols I_A and I_R would distinguish these concepts. I_A, then, is conceptualized as a source of interference resulting from competing ${}_{S}H_{R}$'s (due to interacting stimulus traces), while I_R denotes a fatigue or work-decrement variable.

tulate may be considered plausible and its associated hypothesis tenable. The curves exhibited in Fig. 5 lend additional credence to an interference theory, since the error-savings as functions of rest are maximal for the lowest-$\bar{m}$ list.

Summary and Conclusions

The empirical relationship between performance in serial verbal learning and the attribute of stimulus meaning (*m*) was investigated. Under a standard serial anticipation learning procedure, 72 college students learned three lists of 12 dissyllabic words differing in average meaningfulness ($\bar{m}$), as defined by an association frequency method. The results were as follows:

1. Verbal learning difficulty, measured by the mean trials required to attain one errorless trial, was a decreasing curvilinear function of $\bar{m}$.

2. For three pre-experimental ability levels, the difficulty in attaining a criterion of 7 out of 12 items correct was a decreasing curvilinear function of $\bar{m}$.

3. The introduction of a 2-min. rest pause of color-naming following learning to a criterion of 7 out of 12 items correct resulted in no significant difference in the mean number of items recalled on the first postrest trial (nonreminiscence), but it did yield a nonsignificant work-saving in proceeding to mastery. This finding confirms all studies on reminiscence in which nonsense syllables have not been used.

4. There was a perfect inverse relationship between $\bar{m}$-value and the mean number of serial position errors incurred in attaining a criterion of 7 out of 12 items correct.

5. The introduction of the 2-min. rest at 7/12 mastery resulted in decreased errors for all lists, the error-savings being greatest for the list of lowest $\bar{m}$-value.

6. These findings are interpreted as consistent with a theory of habit interference which predicts that the acquisition of skill in verbal learning is a positive or increasing function of stimulus meaning (*m*).

(Manuscript received August 29, 1951)

References

1. Buxton, C. E. The status of research in reminiscence. *Psychol. Bull.*, 1943, **40**, 313–340.
2. Buxton, C. E. Repetition of two basic experiments on reminiscence in serial verbal learning. *J. exp. Psychol.*, 1949, **39**, 676–682.
3. Buxton, C. E., & Bakan, M. B. Correction vs. non-correction learning techniques as related to reminiscence in serial anticipation learning. *J. exp. Psychol.*, 1949, **39**, 338–341.
4. Buxton, C. E., & Ross, H. V. Relationship between reminiscence and type of learning technique in serial anticipation learning. *J. exp. Psychol.*, 1949, **39**, 41–46.
5. Dysinger, Dale W. A factorial study of speed of response in simple cancellation tasks. Unpublished M.A. thesis, State Univer. Iowa, 1949.
6. Glaze, J. A. The association value of nonsense syllables. *J. genet. Psychol.*, 1928, **35**, 255–269.
7. Hovland, C. I. Experimental studies in rote-learning theory. I. Reminiscence following learning by massed and by distributed practice. *J. exp. Psychol.*, 1938, **22**, 201–224.
8. Hovland, C. I. Experimental studies in rote-learning theory. II. Reminiscence with varying speeds of syllable presentation. *J. exp. Psychol.*, 1938, **22**, 338–353.
9. Hovland, C. I. Experimental studies in rote-learning theory. III. Distribution of practice with varying speeds of syllable presentation. *J. exp. Psychol.*, 1938, **23**, 172–190.
10. Hovland, C. I. Experimental studies in rote-learning theory. VII. Distribution of practice with varying lengths of lists. *J. exp. Psychol.*, 1940, **27**, 271–284.
11. Hovland, C. I., & Kurtz, K. H. Experimental studies in rote-learning theory. IX. Influence of work-decrement factors on verbal learning. *J. exp. Psychol.*, 1951, **42**, 265–272.
12. Hull, C. L. The meaningfulness of 320

selected nonsense syllables. *Amer. J. Psychol.*, 1933, **45**, 730–734.

13. HULL, C. L. *Principles of behavior.* New York: D. Appleton-Century, 1943.
14. HULL, C. L., HOVLAND, C. I., ROSS, R. T., HALL, M., PERKINS, D. T., & FITCH, F. B. *Mathematico-deductive theory of rote learning.* New Haven: Yale Univer. Press, 1940.
15. KREUGER, W. C. F. The relative difficulty of nonsense syllables. *J. exp. Psychol.*, 1934, **17**, 145–153.
16. LINDQUIST, E. F. *Design and analysis of experiments in psychology and education.* Unpublished manuscript, State Univer. of Iowa Library, 1951.
17. LYON, D. O. The relation of length of material to time taken for learning, and the optimum distribution of time. *J. educ. Psychol.*, 1914, **5**, 1–9; 85–91; 155–163.
18. MCGEOCH, J. A. The influence of associative value upon the difficulty of nonsense-syllable lists. *J. genet. Psychol.*, 1930, **37**, 421–426.
19. MCGEOCH, J. A. *The psychology of human learning.* New York: Longmans, Green, 1942.
20. MANN, H. B., & WHITNEY, D. R. On a test of whether one of two random variables is stochastically larger than the other. *Ann. math. Statist.*, 1947, **18**, 50–60.
21. MELTON, A. W. A comparative study of the materials employed in experimental investigations of memory. Unpublished M.A. thesis, Yale Univer., 1929.
22. MELTON, A. W. A comparative study of the materials commonly employed in experimental investigations of memory. Unpublished Ph.D. thesis, Yale Univer., 1932.
23. MELTON, A. W., & STONE, G. R. The retention of serial lists of adjectives over short time-intervals with varying rates of presentation. *J. exp. Psychol.*, 1942, **30**, 295–310.
24. NOBLE, C. E. Absence of reminiscence in the serial rote learning of adjectives. *J. exp. Psychol.*, 1950, **40**, 622–631.
25. NOBLE, C. E. An analysis of meaning. *Psychol. Rev.*, 1952, **59** (in press).
26. SHEFFIELD, F. D. The role of meaningfulness of stimulus and response in verbal learning. Unpublished Ph.D. thesis, Yale Univer., 1946.
27. SHIPLEY, W. C. The effect of a short rest pause on retention in rote series of different lengths. *J. gen. Psychol.*, 1939, **21**, 99–117.
28. STONE, G. R. Retention of nonsense syllables over short intervals of time. *Proc. Ind. Acad. Sci.*, 1946, **55**, 178–181.
29. UNDERWOOD, B. J. *Experimental psychology.* New York: Appleton-Century-Crofts, 1949.
30. WARD, L. B. Reminiscence and rote learning. *Psychol. Monogr.*, 1937, **49**, No. 4 (Whole No. 220).
31. WILSON, J. T. The formation and retention of remote associations in rote learning. *J. exp. Psychol.*, 1949, **39**, 830–838.
32. WITHEY, S., BUXTON, C. E., & ELKIN, A. Control of rest interval activities in experiments on reminiscence in serial verbal learning. *J. exp. Psychol.*, 1949, **39**, 173–176.
33. WITMER, L. R. The association value of three-place consonant syllables. *J. genet. Psychol.*, 1935, **47**, 337–359.
34. WOODWORTH, R. S. *Experimental psychology.* New York: Holt, 1938.

PARTIAL REINFORCEMENT IN A GAMBLING SITUATION[1]

DONALD J. LEWIS

Human Resources Research Office, The George Washington University

In the area of partial reinforcement, there have been relatively few attempts to make experimental tests of the theoretical applications of learning theory to molar human behavior. Humphreys (4) found that the verbal expectations of human *Ss* reacted to partial reinforcement in much the same way as did the "expectations" of lower animals. Grosslight and Child (3) obtained similar findings about the extinction rate of a lever-pulling response. Gilinsky and Stewart (2) found that aspirations that are partially reinforced are more resistant to extinction than those continuously reinforced. Murphy (6), however, did not find the "Skinner effect" in his pinball experiment. These experiments, it should be noted, involved situations that would not be considered important to the *Ss*. The present experiment employed a situation that could be considered a "real" and important one to the human *Ss* used. In this experiment the *Ss* pushed buttons in a "game" with which they were presented. The response of pushing a button was rewarded continuously for some *Ss*, partially for others, and not at all for still others. The rate of extinction of the response was determined with reference to these conditions of learning.

METHOD

Apparatus.—The apparatus was a plywood box, the front face of which supported a red and a blue light. On a small platform extending out from this face toward *S* were four buttons. The apparatus was wired so that the buttons, when pushed by *S*, made a contact that turned on either the red or blue light, but never both. Out of sight of *S* was a small throw switch with which *E* determined whether the red or blue light would go on. The rewards or reinforcements used in the experiment were toys—small plastic cowboys and football players that were purchased from a novelty company.

Experimental design.—The *Ss* were assigned to one of four groups randomly with Groups I, II, and III having 25 *Ss* each and Group IV having 20. Each group was given a different condition of reinforcement. Groups I, II, and III were given a series of 10 acquisition trials, with varying conditions of reinforcement, followed by a series of extinction trials. Group IV was never reinforced, all of its trials being extinction trials from the beginning. The design of the experiment was as follows: "x" signifies reinforcement and "o" signifies no reinforcement.

Group	Reinforcement	Acquisition	Extinction
		1 2 3 4 5 6 7 8 9 10	1 2 3 4 . . . n
I	100%	x x x x x x x x x x	o o o o . . . o
II	50%	x x o o x o x o o x	o o o o . . . o
III	60%	x x x o o o o x x x	o o o o . . . o
IV	0%	o o o o o o o o o o	o o o o . . . o

Subjects.—All *Ss* were between 6.5 and 7.5 years of age. They were boys obtained from primary schools of Inglewood, California. Their IQ's were between 90 and 110. Inglewood is a fairly homogeneous middle-class community and it was assumed that most *Ss* were of comparable economic status. It was requested that no problem children be included among the *Ss*, but this could not be controlled in any other way than on the basis of the teachers' judgments. Some *Ss* were obtained from a summer camp in Inglewood. The same controls were exercised for these *Ss* as for the regular school children, and the summer camp boys were randomly mixed into the four experimental groups so that they could not bias any one group.

Procedure.—The experiment was presented to *S* as a game that he was to play. He was given 20 toys before the first trial and was told that it was possible for him to win many more, but that he also might lose all his toys. The instructions were given orally, and were repeated until *S* indicated that he understood how the game was to be played. Each *S* was tested individually.

[1] This is a part of a dissertation submitted in partial fulfillment of the requirements for the Ph.D. degree at the University of Southern California. The author wishes to express his appreciation to Dr. Milton Metfessel and to Dr. William Grings for their helpful suggestions and encouragement.

He sat immediately in front of the apparatus, which was placed on a table with *E* sitting just to the side.

The instructions to *S* could not be read aloud by *E* from a card to insure standardization. This was found to be impractical because *S* neither listened nor understood. More individual attention had to be given. Still, a strong effort was made to give each *S* the same instructions, and the same attitude toward the game. The *S*s were told:

> First you take a toy out of your bag without looking in the bag. You never look in the bag or feel of it to see how many toys you have left. That would be cheating. You take the first toy you come to out of the bag and put it in this little box alongside of the game. Then you push a button, any button you care to. If the red light goes on, I reach back into this big box of toys I have, and without looking, take the first one I come to, put it in this little box and they are both yours. You take them and put them in your bag. Then you take another toy from your bag, put it in the little box and push a button. Let's pretend that this time the blue light goes on. This time I win and you lose. So I take the toy and put it in the big box and I keep it. Before you push a button you have to put a toy in the little box. You can play as long as you want. You could win a whole bagful of toys, or you could lose them all. But you can quit whenever you want. Just tell me that you want to quit and all the toys in the bag at that time are yours. Are you ready? All right, let's begin.

The *E* entered the game only to see that the game was played properly and to act as "croupier," collecting the lost toys and dispensing the ones *S* won. This also acted to some extent as a control of the time interval between trials. Although *E* did not use a stop watch, he picked up and dispensed the toys at an even rate, trying to keep the intertrial interval as uniform as possible. The interval did vary, however, for a few *S*s took considerable time in deliberation each time before they pushed a button.

Each *S* was requested after he had finished playing not to tell anyone about the game, but undoubtedly some communication did take place. It is not felt, however, that this communication had any significant influence on the results. For one reason, no *S* ever gave evidence of discovering that the game was "rigged," or even of suspecting that it was. At intervals, some *S*s were asked before they started to play if they had any idea about how the game worked. Most of them said "no." A few had such an unspecific idea as "you have to push a button to win a toy." Some *S*s stated an idea about what button, or series of buttons, should be pushed in order to win, but since they all ended with a series of extinction trials, the theory was always proven to be incorrect. A few *S*s, during the extinction series, thought that the machine was broken. When they asked, they were assured that this was not so; that the machine was functioning perfectly.

Motivation for the game was exceptionally high. Just curiosity about the game seemed to arouse considerable motivation. Then, when *S* saw the toys he could win, he became extremely eager to begin. Occasionally, an *S* would say that he did not care for the toys, or that he already had too many of them at home and did not want any more. It was found, however, that in playing the game, these *S*s became just as excited as any of the others. Almost all *S*s made gestures or verbal expressions of annoyance when they lost and appeared very pleased when they won.

Each *S* either won or lost according to the schedule for the experimental group into which he happened to fall. After the fifth extinction trial, *S* was reminded that he could quit when he wanted and that he could keep all the toys that were in the bag at the time of quitting. The play continued either until *S* decided to quit or until all of his toys were gone. Then *S* was asked not to tell anyone about the game. It took from 15 to 20 min. to test each *S*.

Results and Discussion

Table 1 shows the relevant measures of habit strength obtained during extinction. The responses of the continuously reinforced group (Group I) were very significantly more suscep-

TABLE 1

Measures Relevant to Habit Strength as Measured by Extinction

Measure	Group			
	I	II	III	IV
No. toys at beginning of extinction	30	20	22	20
Mdn. no. responses to extinction	8.0	15.0	16.2	10.5
Mdn. no. toys *S* had at time of quitting	22	5	6	10

TABLE 2

χ^2 OF DIFFERENCES BETWEEN GROUPS*

Groups	χ^2	Groups	χ^2
I & II	14.79	II & III	0.00
I & III	11.52	II & IV	3.93
I & IV	2.37	III & IV	7.17

* χ^2 of 3.84 significant at .05 level of confidence; χ^2 of 6.64 significant at .01 level of confidence.

tible to extinction than those of either of the partially reinforced groups (Groups II and III). These differences are shown in Table 2.

The median number of responses to extinction of the *Ss* of Group IV, who never received reinforcement, was 10.5. For the continuously reinforced Group I, the median was 8.0, the difference being below that demanded for significance at the 5% level. This probably means that the response had a fairly high habit strength, for all groups, before a button was pushed. Undoubtedly previous button-pushing experience generalized to this situation. It should be noted that many of the partial reinforcement studies on the maze-running habit in the rat begin the partially reinforced series only after the response has acquired considerable habit strength due to continuous reinforcement. It can probably be concluded, for the present experiment, that the ten reinforcements that Group I received did not make the response perceptibly stronger, in terms of resistance to extinction, than it was before reinforcements were given. Even though the actual number of reinforcements that Groups II and III received was fewer than received by Group I, the habit strength of the partially reinforced groups was greater. The number, or amount, of reinforcements, then, would seem to be irrelevant in this situation. It seems that the pattern in which the reinforcements were administered was probably the significant variable.

Sheffield (7) has indicated how this patterning of reinforcement achieves its effect. The superior resistance to extinction of Groups II and III of this experiment can be explained by pointing out that the button-pushing responses for both of these groups have as stimuli, before the reinforcing trials are completed, the aftereffects of both success and failure. If on one trial, for example, the button-pushing response is successful, the aftereffects of this reinforcement remain until the next trial and become part of the total stimulus for this next trial. The succeeding trial then may bring nonreinforcement. The aftereffects of this trial become part of the total stimulus for the third trial. And so on through the partially reinforced series. The continuous reinforcement series, however, is relatively homogeneous. The stimulus pattern changes very little from trial to trial. Thus, when the extinction series starts, the responses of the continuous group will be weaker due to stimulus generalization than will the responses of the partially reinforced groups. This is because the extinction stimuli are markedly different from the reinforcement stimuli for the continuous group. For the partial groups, the stimuli do not differ as much from the reinforcement series to the extinction series.

Perhaps among the more important of the aftereffects of a nonreinforced trial is frustration. Finger (1) has suggested that a considerable amount of frustration would be engendered by the onset of extinction. If that is true, a partially reinforced group would already have become adapted to the frustration; frustration would in fact serve as a stimulus for succeeding reinforcement or success. This would not be true for a continuously reinforced group and, possibly, the frustration for this group would lead to an emotional blocking of the response and so lead fairly quickly to extinction. Skinner (8) has found that the extinction curve after partial reinforcement is relatively smooth, but that the extinction curve following continuous reinforcement is characterized by cyclic emotional fluctuations. Jenkins, McFann, and Clayton (5) have found this same post-continuous cyclic fluctuation. Perhaps what the clinicians refer to when they talk about "frustration tolerance" is very similar to the effects of partial reinforcement. It is suggested that the present experimental setup would provide a situation in which frustration tolerance could be studied experimentally.

No significant difference was found between Groups II and III, the two

partially reinforced groups. This was true even though Group II was reinforced 50% of the time and Group III 60% of the time. This 10% difference in the amount of reinforcement, and in the distribution of the reinforcement, had no significant effect upon the number of responses that were made during extinction.

Since not all the *Ss* used in this experiment had the same number of toys in their possession at the beginning of extinction, it could be conjectured that *S* was guided by the number of toys he had left in his bag when making his decision to quit. This was controlled in part by instructing *S* never to look into or feel of the bag. Furthermore, the *Ss* in Groups I and IV had *more* toys at the time of quitting than the *Ss* in Groups II and III. It seems safe to conclude that the different conditions of reinforcement were at least the sufficient conditions for the different extinction rates.

Summary

The present experiment was designed to test the adequacy of the notion of partial reinforcement for a situation that could be considered a real and important one to the human *Ss* used. The *S* was presented with a "gambling" game in which he could win or lose small toys. Four groups were used: during ten acquisition trials, one group received continuous reinforcement, one received 50% partial reinforcement, and one received 60% partial reinforcement. The fourth group was never reinforced at any time. The three groups that had received reinforcement then underwent extinction trials. It was found that the partially reinforced groups were very significantly more resistant to extinction in this situation than was the continuously reinforced group. The group that had never received reinforcement was not significantly different in resistance to extinction from the group that had been continuously reinforced during acquisition. A reinforcement theory interpretation of the results was made, and it was decided that "partial reinforcement" seemed to have a considerable degree of explanatory power on this molar level.

(Manuscript received September 4, 1951)

References

1. Finger, F. W. Retention and subsequent extinction of a simple running response following varying conditions of reinforcement. *J. exp. Psychol.*, 1942, **31**, 120–134.
2. Gilinsky, A. S., & Stewart, J. C. Extinction of a success operation following three conditions of reinforcement. *Amer. Psychologist*, 1949, **4**, 222–223. (Abstract)
3. Grosslight, J. H., & Child, I. Persistence as a function of previous experience of failure followed by success. *Amer. J. Psychol.*, 1947, **60**, 378–387.
4. Humphreys, L. G. Acquisition and extinction of verbal expectations in a situation analogous to conditioning. *J. exp. Psychol.*, 1939, **25**, 294–301.
5. Jenkins, W. O., McFann, H., & Clayton, F. L. A methodological study of extinction following a periodic and continuous reinforcement. *J. comp. physiol. Psychol.*, 1950, **43**, 155–167.
6. Murphy, M. J. The differentiation of a response: An exploratory study with human subjects. Unpublished Ph.D. dissertation, New York Univer., 1943.
7. Sheffield, V. F. Extinction as a function of partial reinforcement and distribution of practice. *J. exp. Psychol.*, 1949, **39**, 511–525.
8. Skinner, B. F. *The behavior of organisms.* New York: D. Appleton-Century, 1938.

JOURNAL OF EXPERIMENTAL PSYCHOLOGY

The issues listed below are available for sale

Year	Volume	Available Numbers						Price per Number	Price per Volume
1916	1	1	2	3	4	5	6	$1.25	$7.00
1917	2	1	2	–	4	5	–	$1.25	$5.00
1920	3	–	2	–	4	5	–	$1.25	$3.75
1921	4	1	–	3	4	5	6	$1.25	$6.25
1922	5	–	–	3	4	5	–	$1.25	$3.75
1923	6	–	–	–	–	–	–	——	——
1924	7	1	2	–	–	5	–	$1.25	$3.75
1925	8	1	2	3	4	–	6	$1.25	$6.25
1926	9	–	2	3	–	5	6	$1.25	$5.00
1927	10	–	–	3	4	5	6	$1.25	$5.00
1928	11	1	2	3	4	5	6	$1.25	$7.00
1929	12	1	2	3	4	5	6	$1.25	$7.00
1930	13	1	2	3	4	5	6	$1.25	$7.00
1931	14	1	2	3	4	5	6	$1.25	$7.00
1932	15	1	2	3	4	5	6	$1.25	$7.00
1933	16	1	2	3	4	5	6	$1.25	$7.00
1934	17	1	2	3	4	5	6	$1.25	$7.00
1935	18	1	2	3	4	5	6	$1.25	$7.00
1936	19	1	2	3	4	5	6	$1.25	$7.00
1937	20	1	2	3	4	5	6	$1.25	$7.00
	21	1	2	3	4	5	6	$1.25	$7.00
1938	22	1	2	3	4	5	6	$1.25	$7.00
	23	1	2	3	4	5	6	$1.25	$7.00
1939	24	1	2	3	4	5	6	$1.25	$7.00
	25	1	2	3	4	5	6	$1.25	$7.00
1940	26	1	2	3	4	5	6	$1.25	$7.00
	27	1	2	3	4	5	–	$1.25	$6.25
1941	28	1	2	3	4	5	6	$1.25	$7.00
	29	1	2	3	4	5	6	$1.25	$7.00
1942	30	1	2	3	4	5	6	$1.25	$7.00
	31	1	2	3	4	5	6	$1.25	$7.00
1943	32	1	2	3	4	5	6	$1.25	$7.00
	33	1	2	3	4	5	6	$1.25	$7.00
1944	34	1	2	3	4	5	6	$1.25	$7.00
1945	35	1	2	3	4	5	6	$1.25	$7.00
1946	36	1	2	3	4	5	6	$1.25	$7.00
1947	37	1	2	3	4	5	6	$1.25	$7.00
1948	38	1	2	3	4	5	6	$1.25	$7.00
1949	39	1	2	3	4	5	6	$1.25	$7.00
1950	40	1	2	3	4	5	6	$1.25	$7.00
1951	41	1	2	3	4	5	6	$1.25	$7.00
	42	1	2	3	4	5	6	$1.25	$7.00
1952	43	By subscription, $14.00 (2 vols.)						$1.25	——
	44								

List price, Volumes 1 through 42 $269.00
30% discount 80.70

Net price, Volumes 1 through 42 $188.30

Information about the Journal of Experimental Psychology: as noted in the table, the journal was not published in 1918 and 1919; from 1937 to 1943 it was published at the rate of two volumes per year; from 1944 to 1950, in one volume per year. Beginning in 1951 there are two volumes a year.

Information about prices: when all numbers are available, the price per volume throughout the series is $7.00. For incomplete volumes, the price is $1.25 for each available number. For foreign postage, $.25 per volume should be added. The American Psychological Association gives the following discounts on orders for any one journal:

10% on orders of $ 50.00 and over
20% on orders of $100.00 and over
30% on orders of $150.00 and over

Address orders and subscriptions to

AMERICAN PSYCHOLOGICAL ASSOCIATION

1515 Massachusetts Avenue N.W. Washington 5, D. C.

M 1134